OCCUPATIONAL AND EARNING DYNAMICS, THE ROLES OF WEALTH AND EDUCATION IN THE RURAL NON-FARM ECONOMY: EVIDENCE FROM THAILAND AND INDONESIA

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Chayanee Chawanote

January 2014

OCCUPATIONAL AND EARNING DYNAMICS, THE ROLES OF WEALTH AND EDUCATION IN THE RURAL NON-FARM ECONOMY: EVIDENCE FROM THAILAND AND INDONESIA

Chayanee Chawanote, Ph. D.

Cornell University 2014

The first paper, "Farm and Non-farm Occupational and Earnings Dynamics in Rural Thailand", explores individual occupational and earnings dynamics in rural Thailand during 2005-2010. We find significant occupational transitions, mainly involving moving out of farming and into non-farm employment, rather than starting businesses, especially enterprises that employ others. Using stochastic dominance, we identify an occupational ladder, with the most remunerative employment as a non-farm business owner/employer, and the worst as an agricultural worker. Occupational transitions into the rural non-farm economy are associated with statistically significant earnings gains while transitions into farming are associated with earnings losses. These results are confirmed with a variety of methods to control for prospective unobserved heterogeneity. However, a small number of individuals become and remain non-farm employers, reflecting the difficulty in operating non-farm businesses that employ others.

The second paper, "The Correlates and Dynamics of Rural Household Non-farm Business and Entrepreneurial Job Creation in Thailand", explores the characteristics of those non-farm household entrepreneurs who expand their businesses by hiring non-family members, as well as the push and pull factors that play a role in supporting job creation within the rural non-farm economy and the

dynamics of rural non-farm business status. More than 90 percent of rural non-farm enterprises are self-employed without non-family workers and rarely grow larger into enterprises that provide jobs outside the family. Instead, we observe a tendency for microenterprises and small-medium enterprises to contract in size. Non-farm businesses who hire more workers have more household assets, smaller agricultural land holdings, and lower opportunity costs of labor related to other household activities. Credit availability also plays an important role in starting an enterprise and hiring more workers.

The third paper, “Intersectoral and Gender Heterogeneity in the Marginal Earnings Gains Associated with Education in Indonesia”, studies how individual earnings respond to added educational attainment given intersectoral labor market heterogeneity and gender differences, as well as the potential signaling that comes from educational credentialing in the job market. We focus on the problem of non-random selection into sector of employment. Using the Indonesian Family Life Survey data from 1993-2007, we use parents’ occupation and education as instruments influencing individuals’ sector of employment. We find evidence of both sheepskin effects associated with the completion of specific, multi-year stages of schooling and intersectoral differences in the estimated marginal earnings gains associated with educational attainment, all of which vary between females and males. These findings carry significant implications for understanding patterns of private investment in education if not all children are equally likely to enter all careers and marginal earnings gains associated with educational attainment vary markedly among future work patterns.

BIOGRAPHICAL SKETCH

Chayanee Chawanote earned her Bachelors degree in Economics (with first class honors) from Thammasat University in 2005. She also received her Masters degree in Economics (English program) from Thammasat University in 2008. In 2006, she started working as a lecturer at the Faculty of Economics, Thammasat University and in 2008 received a Thammasat scholarship to pursue her Doctoral degree at Cornell University.

I dedicate this to my mother and father.

Thank you for unconditional love and support for your only child and taking care of yourselves while I was away in the U.S. I really miss and love you.

ACKNOWLEDGMENTS

"Life is a journey" is one of my favorite quotations that gives me encouragement to move forward to pursue my dreams, and break through all obstacles, big or small. Studying for a Doctoral degree is one of my top dreams for as long as I can remember, even though I had no clue as to what field I should aim at for the Ph.D.

Above all, I would like to thank my parents for their unequivocal support throughout, as always, for which my mere expression of thanks does not suffice. Being high school teachers, my parents guided me to pay careful attention in Mathematics and Social Sciences that later on make me become a person with logical senses who always has eyes open to the social issues affecting our life and the majority of the people. With my great interest in both fields, I matriculated ato Thammasat University, majoring in Economics, under the oldest Faculty of Economics in Thailand.

Studying at Thammasat University helped me frame my view in academic career gave me chances to meet many incredible mentors who became my role models in both research integrity and teaching philosophy. Ajarn Pranee Tinakorn, my first role model, who impressed me with her econometrics class and mentioned that econometrics is both art and sciences. Ajarn Rungsan Thanapornpan, who hired me to be one of his research assistants right after I finished my undergraduate degree. He gave such invaluable academic experiences including my first academic article. Ajarn Patamawadee Pochanukul, who encouraged me to pay a close attention to rural development and development economics, although I have never enrolled in her class. I am most grateful to Ajarn Jakrapong Uchupalanan who was my adviser during my Masters degree at Thammasat University. He is also a Cornell alumnus who suggested to me to apply for a Ph.D. at the Department of Applied Economics and Management. I also would like to thank all the instructors who are always supportive when I need some advice and always make me feel at home whenever I made a visit to the Faculty of Economics during my Ph.D. studiers. Most importantly, I would like to acknowledge the financial support from Thammasat University.

Coming to study Ph.D. at Cornell is one of the most noteworthy journeys of my life. The education system in the US forces me to improve my critical thinking skills. Cornell really does prove its motto of “I would found an institution where any person can find instruction in any study.”

It would not have been possible with for me to go through all the academic difficulties without the help, support and patience of my principal supervisor, Professor Chris Barrett, not to mention his advice and unsurpassed knowledge of professional training toward the research frontier, and feedback on my work. He is a great mentor who dedicates his time to his students, puts up with my slowing on the uptake, and be encouraging. I am taking Chris as my role model in by all means especially in mentoring style. I know I cannot thank him enough for what he has helped me during my studies at Cornell.

Many other economists played important roles in my graduate training, for which I am very thankful: all professors I took classes with; my dissertation committee: Professor Gary Fields who emphasized how to write a good research paper, asking the right questions and answer them clearly, Professor Garrick Blalock who provided some tricks on how to manage data and that helped me working with my data very conveniently; other professors who helped me develop my dissertation at some points either for econometrics issues or interesting discussions: Professor George Jakubson, Professor Jim Berry, Professor Damon Clark, Professor Francesca Molinari, and other professors who came to give a seminar at Cornell.

I would like to have special thanks to my colleagues: Pin Sommarat who always answered my questions even though she is working in another continent and very busy, David Newhouse who is my co-author in another paper using IFLS data and helped me tremendously in understanding how to use IFLS data, Brian Dillon who is my co-author in chapter 4 and patiently help me develop, answer my questions, and revise a paper, Russell Toth who is always supportive and provided me all useful literature, insightful discussions and information, and my classmates who tried to answer many questions I have: Kira Maria Villa, Elaine Hill, Xi Chen, Yu Qin, Catalina Herrera, Priya Mukherjee, and Xin Jin.

In addition to those already listed, many other people provided helpful feedbacks and very useful suggestions on the work in this dissertation, whether at conferences and Cornell seminars, in conversation, or by reading drafts. These include in no particular order: Sosina Bezu, Paul Cichello, Steve Haggblade, Peter Hazell, Julia Berazneva, Joanna Upton, Aurelie Harou, Annemie Maertens, Marc Rockmore, Tom Walker, Teevrat Grag, Nathan Jensen, Leah Bevis, Andrew Simons, Ellen McCullough, Jennifer Cisse, and Kibrom Hirfrfot.

There are many people helping me to obtain Thailand data used in the dissertation. I appreciate Wittawat Hemtanon at Prince of Songkla University who helped me conduct a small field survey and accommodated me during my visit in Had Yai, Ajarn Supachai Srisuchart at Thammasat University who introduced me to the Thai SES panel data and a contact person at the National Statistical Office (NSO) of Thailand, Ajarn Nat Puttanapong who provided me with Thailand rainfall data. Not to mention the Cornell Graduate School Research Travel Grant that supported this dissertation. I also would like to thank the NSO and the University of Chicago-UTCC Research Center for giving me a permission to use their data.

There are times I felt really depressed when things did not go as well as I planned and times when I thought I should just give up and felt lost in this academic world. Besides my classmates, for international students like me, friends from the same country often share similar difficulties. There are many Thai senior students at Cornell I would like to thank especially for helping me on both academic issues and living in the U.S. issues: P'Khwan who was my first housemate and my classmate during our first year, living and experiencing hardship with me for 4 years, P'Yee who was my generous brother in Ithaca and even when he moved to New York City, P'Teelek who was another big brother. P'Nij and P'Aeh who helped me with a GIS program and some academic writing reviews, P'Pooky and P'Fresh who were also my AEM seniors and classmates when we did homework in classes we took together, P'Golf who is studying at University of Wisconsin at Madison but always ready to chat and answer my questions. In addition, I am very grateful to my same-age friends and junior friends: Pimbucha who has become my close friend, discussing on our

research and almost everything in our life (I already lost count of how many trips we traveled together), Sunsiree who was my last housemate and has become like my younger sister, Annie who established herself to be a president of Namtip fan club and cooked meals during my busy weeks writing a dissertation for the B exam (and made me teary eyed), lovely friends: Kunlaya, Nongnuch, Jeeraporn, Sra, Natt, and Passaporn who are always beside me and helped make my life easy while studying at Cornell, and other Thai friends I did not mention. We spent many great memories while living in Ithaca.

Amongst my supportive Masters degree friends in the Faculty of Economics, the effort made by Aim whom always exchanged our diary via email from time to time. One thing we have learned from our idols we are fan clubs is that they never give up and work really hard to achieve where they are and that is applied to everyone who wants to follow their dream. I really think that our email exchanges can be published into a (self-made) diary essay of our life journey.

Lastly, but by no means least, again, I would like to thank my parents who provided constant love, and have been unwavering in their support and listening to any of my difficulties throughout my stressful years of graduate school. Just hearing their voices already made me feel relieved.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

In the process of structural transformation, the rural non-farm economy (RNFE) serves as a key linkage between the agricultural sector and the macroeconomy. Lewis (1954) saw the non-farm sector as potentially dynamic and expanding. In his view and in the dual economy models that followed from his seminal work, the RNFE can absorb a surplus farm labor supply from rural households to mix with industrial capital, providing more productive use of labor and hence earning higher returns and providing an engine for economic growth. With agricultural transformation, higher labor productivity in agriculture allows households to allocate more time to non-farm activities or to migrate to the urban industrial sector (Timmer, 2009). Anticipating this, rural households invest their profits from agriculture in human capital formation, raising non-farm labor productivity. All of these mechanisms lead to a growing rural non-farm sector and gradual shrinkage in the intersectoral differences in the marginal returns to labor (Barrett et al., 2010). However, the process of integrating farm labor into the non-farm economy and equalizing factor returns across sectors takes a long time as it requires well-functioning labor and capital markets throughout an economy. Hence, we observe intersectoral differences in earnings and returns, as well as in productivity in most developing countries.

The RNFE is increasingly seen as a pathway out of poverty in low- and middle-income countries (Timmer 1988, 2002). The benefits from agricultural transformation and productivity growth are not distributed equally to all farmers and labors in agricultural sector. In particular, many land-constrained rural households do not reap these benefits and are too often left out of the agriculture-led economic growth process. Their earnings fall behind the rest of the economy. That is when the RNFE comes to play a crucial role because its heterogeneity in size and activities may accommodate rural households from various backgrounds and endowments, including those that might otherwise be left out of agriculture-led growth. Most rural households earn at least some income from non-farm sources, as non-farm workers or owner/operators of non-farm businesses, or both (Reardon 1997, Foster 2012). Rural non-farm incomes account for an estimated 35-50 percent of rural incomes across the developing world, including 51% in Asia during 1990s and 2000s (Haggblade et al., 2007).

Whether the RNFE has helped with poverty alleviation and promoting upward mobility remains debatable, however. The heterogeneity of RNFE activities implies there are high- and low-return activities in which different rural households can engage. Oftentimes, poor households appear pushed out of the agricultural sector or their traditional occupation into low-return non-farm activities with low capital startup requirements in highly competitive sub-sectors or just part-time employment. In contrast, high-return non-farm activities often require significant initial capital and skilled labor often possessed only by better-off households that can respond to the pull of profitable non-farm opportunities and thereby enjoy greater earning gains than do participants pushed into the RNFE (Barrett et al. 2001). When we focus on productive household assets and specific factors of production such as human, physical, and financial capital stocks, the poor with low capital stocks often face barriers to entry or have limited involvement in more remunerative non-farm activities (Barrett 1997, Dercon and Krishnan 2000, Barrett et al. 2001, Haggblade et al. 2007).

The dynamic role of the RNFE and its role in promoting upward earnings mobility thus remains poorly understood. This dissertation investigates further occupational choices across agricultural and non-agricultural sectors and their associated earnings dynamics (Chapter 2, "Farm and Non-farm Occupational and Earnings Dynamics in Rural Thailand"). Little is known about earnings and occupational mobility in rural Thailand. We are also able to distinguish individuals' occupational status as worker, self-employed, and employers. With six distinct groups–each of the three occupational groups in the farm and non-farm sectors–in nationally representative survey data over a period of five years, we can paint a far richer picture of rural occupational transitions and earnings dynamics than exists in the prior literature.

We find that transitions mainly involve moves into farm self-employment or non-farm employment rather than into non-farm self-employment, much less into farm or non-farm employer positions. People in the farm sector tend to move within their sector and some transition to non-farm employee while some of non-farm sector move into farm self-employment. The non-farm employers' and employees' earnings distributions stochastically dominate the other categories' earnings distributions, while those associated with farm workers and self-employed farmers are stochastically dominated by each of the three non-farm occupational groupings. All these results strongly suggest a "gravity effect" on the occupation ladder. It is easier to move down into lower-return occupations, especially non-farm self-employment and self-employed farming, than up into higher-return ones like non-farm business employer and formal non-farm employee. However, a small number of individuals become and remain non-farm employers, the most lucrative occupation, reflecting the difficulty in operating non-farm businesses that employ others.

We then move to multivariate regression analysis to explore how occupational transitions affect directional earnings mobility, employing a conditional mobility model in which change in earnings or change in log earnings (annualized) are regressed on time-invariant and time-varying individual characteristics, using 2005-6, 2006-7, and 2007-10 transitions. We use both individual fixed effects to control for unobserved time invariant characteristics and a Hausman- Taylor estimator, an instrumental variables approach that enables estimation of the coefficients of time-invariant regressors while still controlling for individual-level random effects. Then, we test the differences on these coefficients representing gains (losses) from transitions. We combine all occupations within the farm sector together due to the small number of observations and their joint stochastic dominance by the non-farm options, leaving four occupation categories: farm, non-farm self-employed, non-farm employee and non-farm employer. On average, movements into the non-farm sector increase earnings relative to remaining in the farm sector. By contrast, shifting between non-farm sectors could result in either earnings gains or losses. In most cases of these mixed outcomes, transitions from any of the non-farm occupations into another non-farm position lead to lower earnings growth than does staying in the original position.

Time-varying unobservables might affect both earnings dynamics and occupational transitions. So we refine the estimation method to predict the probability of occupational movements in a first stage and then use these predicted transition probabilities in two-stage estimation of the characteristics most strongly associated with earnings growth. Our instruments are changes in village characteristics, reflecting changes in infrastructure and agricultural circumstances that affect the occupation decisions but are exogenous to individual earnings changes. This identification strategy allows for more variation across villages while still allowing for variation in individual characteristics within villages. As another robustness check, we estimate a multinomial logit model correcting for selection bias, following Dubin and McFadden (1984) and Bourguignon et al. (2007), to account for the fact that occupational changes might be subject to both selection bias and endogeneity. The results of these robustness checks are consistent with the individual fixed effects and Hausman-Taylor estimators. In all of these estimations of earnings changes on occupational transitions and other controls, we consistently find that both higher individual educational attainment and greater household physical capital endowments strongly and statistically significantly increase earnings, indicating the joint importance of physical and human capital, as well as climbing the occupational ladder, to earnings mobility.

These findings motivate further investigation into the roles of physical and human capital in Chapters 3 and 4. Chapter 3 ("The Correlates and Dynamics of Rural Household Non-farm Business and Entrepreneurial Job Creation in Thailand") explores the correlates and dynamics of rural non-farm businesses in Thailand. This chapter focuses in particular on the empirical regularity that we rarely see firm growth from self-employment without non-family employees into a larger firm that provides RNFE jobs outside the family. We explore the role of household assets as crucial to entry into and mobility within the RNFE sector.

Previous studies decompose rural non-farm employment into low- and high-productivity salaried work or self-employment (e.g., Barrett et al., 2005; Jonasson and Helfand, 2009; Bezu et. al., 2012). In this chapter, however, entrepreneurship status is classified separately by its RNFE employment generation effects, from self-employment without employees–what are sometimes referred to as "subsistence" entrepreneurs–to those "transformational" entrepreneurs who create jobs for non-family members and are thus more effective in stimulating broad-based economic growth (Schoar, 2010).

We find that subsistence self-employed businesses do not automatically transition into transformational entrepreneurs. This leads to an oft-observed 'missing middle' in the enterprise size distribution in developing countries, wherein the overwhelming majority of RNFE enterprises are very small, with no paid employees, and a very small number of large employers account for most paid employment, with very few small-to-medium size enterprises (Mead and Liedholm, 1998; de Mel et al., 2008; Mondragon-Velez and Pena-Parga, 2010). In our study, more than 90 percent of rural non-farm enterprises are just self-employed without non-family workers, and these rarely grew larger over the 2005-2010 period we study. Indeed, we observe a tendency to contract in firm size for microenterprises and small-medium enterprises (SMEs), while only three percent graduate to a larger firm size.

The short period of the panel data limits our ability to make strong causal inferences on household assets and wealth in determining rural non-farm business participation. We thus use predetermined household attributes in 2005 and exogenous variation from rainfall and sub-district data to address likely endogeneity issues that necessarily complicate inference from the observed associations between household attributes and enterprise creation and growth performance. The predicted probabilities from a simple linear probability model of ownership of a rural non-farm business in 2010 show interesting heterogeneity in the association with wealth across the household wealth distribution. There is a positive relationship between asset index and the predicted probability of operating a non-farm business across the asset index distribution, but there is a negative relationship between wealth (comprised of more liquid assets) and the predicted probability at the lower range of the wealth distribution. The positive asset relationship supports the hypothesis that liquidity constraints limit rural households from entering into self-employed non-farm businesses as a low return safety net, as well as households who need investible capital to become transformational entrepreneurs in the face of lumpy initial capital requirements in high-return sub-sectors.

Besides confirming the heterogeneous wealth effects on RNFE business ownership, the chapter also provides more insights into other push and pull factors associated with the likelihood of households running a non-farm business or hiring more workers in rural Thailand. An ordered probit model is then used to estimate household choice among four rural household non-farm business statuses in 2010, including no non-farm business, self-employed non-farm businesses without paid employees, microenterprises with one to ten paid employees, and small-medium sized business with more than ten employees. We find that households who run a non-farm business or hire more workers have greater household assets and smaller agricultural land holdings. As predicted in the analytical framework, they allocate less household labor to operating non-farm business when the average non-farm wage within community increases, reflecting an increase in the opportunity cost of household resource allocation as well as in the cost of hiring workers. Credit availability also plays an important role in starting an enterprise and hiring more workers. Excess rainfall discourages households from starting non-farm businesses, suggesting that rural non-farm business creation might not be a good income diversification strategy in Thailand, at least as compared to working in the (skilled) non-farm wage sector if they are able to choose both sectors.

The last part of the chapter explores the dynamics of rural non-farm enterprises using multinomial logit and bivariate probit models. We therefore have four alternative values for the dependent variable: never had a rural non-farm business, entry into, exit from, or remaining in rural non-farm business from 2006 to 2010, with the same set of left hand sided variables as in the previous models. The push and pull factors show a similar pattern as did the results with respect to the probit estimation of operating a rural non-farm enterprise. A college degree is associated with a greater likelihood that a household never owns a non-farm business, signaling a widespread preference for employment over entrepreneurship. Households that previously allocated labor to farming or non-farm employment are likely to remain in that occupation, indicating occupational persistence.

We learn that household assets and capital access are vitally important to being able to participate as a business owner in the RNFE, and especially to growing rural non-farm businesses. On the other hand, greater educational attainment is important to transitions into the (formal) non-farm wage employment sector, rather than into self-employment, similar to findings in India from Kijima and Lanjouw (2005). So human capital formation through formal education seems to militate against RNFE enterprise creation and growth while financial/physical capital is crucial to that process.

Chapter 4 ("Intersectoral and Gender Heterogeneity in the Marginal Earnings Gains Associated with Education in Indonesia") then studies how educational attainment is associated with (i) the probability of employment in the more remunerative non-farm sector, (ii) the estimated Mincerian marginal earnings from additional education in both farm and non-farm work, and (iii) how much of those estimated earnings differentials arise due to education's role in signaling employee characteristics to prospective employers within each sector. By focusing on the Mincerian returns to education, we assume that the cost of education is only the earnings foregone while in school, or that the direct costs and unmeasured earnings while in school cancel out each other (Fields, 1980). This is an entirely descriptive exercise as we cannot fully control for prospective endogeneity and unobserved heterogeneity. The results are nonetheless powerful in revealing important patterns in the relationship between education and earnings in rural areas of a middle-income country.

Heterogeneity in earnings differentials associated with educational attainment may arise because of intersectoral differences in the marginal returns to labor quality, as reflected in school completion. The agricultural household will allocate more labor to non-farm work if the marginal product of human capital is relatively high in non-farm wage work (Huffman and Orazem 2007). But again, household ownership of fixed, sector-specific factors of production, like farmland or farming skills, combined with spatially heterogeneous access to public schools, transportation and communications infrastructure that affect labor productivity could impede or facilitate rural individuals' ability and willingness to invest in education and to reallocate labor intersectorally if the access to these facilities is not identically distributed. The result can be persistent differences both in the likelihood of adult employment in different sectors and in the earnings gains from education across sectors. These joint differences remain largely unexplained in the literature, however.

The econometric challenge arises because sector and earnings are jointly determined by educational attainment. Duflo (2001), Oyelere (2010), Comola and Mello (2013) present evidence, however, that the estimated returns to education are similar in both OLS and IV estimates. Given this repeated finding in the prior literature, endogenous schooling choices might not be a significant source of bias. We therefore focus on the problem of non-random selection into sector of employment. We use parents' occupation and education from at least one decade prior as instruments determining individuals' sector of employment. The underlying idea is that intergenerational employment networks and distant past sector-specific capital (such as farmland) influence one's sector of employment but conditional on sector do not directly affect earnings (Magruder, 2010; Hellerstein and Morrill, 2010; Kramarz and Skans, 2011). Not many studies using this type of information in determining sector of employment that determine earnings (Krishnan, 1996).

Ultimately, our findings emphasize that education plays a key role in job market sorting among different sectors, as well as in explaining earnings differentials across sectors and between men and women. Parental occupation and education jointly influence adult child's current occupation, providing us with a credible way to control for selection into sector of occupation and thereby resolve an important source of prospective bias in the estimated marginal earnings gains associated with education. Once we correct for selection effects, we find clear evidence of both sheepskin effects associated with the completion of specific, multi-year stages of schooling and intersectoral differences in the estimated marginal earnings gains from schooling, all of which vary between females and males. These gender and intersectoral differences arise from both differences in labor market signaling and in the marginal earnings gains related to productivity of labor. The marginal earnings gains from schooling are, on average, highest for non-farm employees. These findings carry significant implications for how we explore parents' optimal educational investments in their children if not all children are equally likely to enter all careers and the marginal earnings gains associated with education vary markedly among future work patterns.

The countries of interest in this study are Thailand and Indonesia. Chapters 2 and 3 focus on the Thai rural economy using the nationally representative Socio-Economic Survey (SES) panel collected annually between 2005 and 2007, and the subsequent round in early 2010. The SES data enable us to use individual observations over a longer period, and in a large-scale nationally representative survey, than do most prior RNFE studies. As a result, we can apply multiple empirical approaches for robustness checks that need more than two periods of observation per individual. We can also use initial year attributes as initial controls and pre-determined covariates for regressors that are likely to be endogenous or co-determined with the dependent variable(s). More importantly, the data enable us to differentiate among alternative farm and non-farm occupations in a way that sheds light into rural development policy debates over whether to invest in promotinge rural household entrepreneurship as a means of stimulating business and employment creation. We also merge the household and individual level survey data with another, village-level dataset, the National Rural Development (NRD) census data set, to give more information on the conditions and infrastructure of the village and local economy. Although the Thai SES data attrition rate is not high, since our study follows only employed individuals and households in all four-survey waves, we pay considerable attention to possible biases introduced due to attrition, estimating and applying the inverse probability weights suggested by Fitzgerald et al. (1998) and Wooldridge (2002).

The last chapter uses the Indonesian Family Life Survey (IFLS) in order to tap the richness of that 1993 to 2007 four wave panel data set. The IFLS enables us to match information on parents' occupation and education, so as to generate credible instruments for the sector in which adult children work and to explore intergenerational effects on current individuals' employment patterns. These controls enable us to address the likely endogeneity of individuals' sector of employment so as to generate less biased estimates of the relationship between educational attainment and individual earnings across sectors and between the sexes, as discussed above.

This dissertation represents an attempt to understand the dynamic role of the

RNFE on earnings mobility and occupational choice, and the prospectively heterogeneous role of human capital (especially that due to education) and financial capital on entry into, growth in, and earnings from the RNFE. We contribute to the literature on the RNFE, rural entrepreneurship in developing countries, and labor economics. Results from the dissertation provide some interesting policy implications, which are discussed in each chapter. Of particular note are the following implications. There exists a striking mismatch between the jobs created by rural household enterprises, which represent less than one percent of household-owned enterprises in rural Thailand, and the jobs held by those who work for a salary or wages in the rural non-farm economy, who comprise a majority of the population. Our findings suggest that promoting rural non-farm employment by larger enterprises may be more important to rural income growth than promoting rural non-farm self-employment and household entrepreneurial activity that might not offer an adequately broad platform to facilitate agrarian transformation and rural earnings growth. Especially if promoting rural non-farm employment is a more expensive drain on the government budget than is encouraging larger, established private firms to establish new locations and jobs in rural areas. In addition, as we observe inter-sectoral difference in wages due to schooling, this has significant implications in optimal educational investments for parents who have different expectations in their children's potential sector of employment. It also provides more understanding why we observed low educational investments when the estimated private returns to education in the literature are routinely high in low-income and middle-income countries. Although the study does not contribute to social returns to schooling in the way that the government use the social rate in deciding their policy to implement, it at least helps in shaping the policy on how to promote educational attainment in respond to wage differentials in economies with imperfect intersectoral labor market integration.

REFERENCES

Barrett, C. B., M. Bezuneh, D. C. Clay, & T. Reardon. (2005). "Heterogeneous constraints, incentives and income diversification strategies in rural Africa." *Quarterly Journal of International Agriculture, 44*(1), 37-60.

Barrett, C. B., T. Reardon, & P. Webb. (2001). "Non-farm income diversification and household livelihood strategies in rural Africa: concepts, dynamics, and policy implications." *Food Policy, 26*, 315-331.

Barrett, C.B., M.R. Carter, and P. Timmer. (2010). "A Century-Long Perspective on Agricultural Development", *American Journal of Agricultural Economics* 92(2): 447-468.

Bezu, S., C.B. Barrett & S. Holden. (2012)."Does Non-farm economy offer pathways for upward mobility? Evidence from a panel data study in Ethiopia," *World Development, 40*(8), 1634-1646.

Bourguignon, F., M. Fournier, and M. Gurgand. (2007). "Selection bias corrections based on the multinomial logit model: Monte Carlo comparisons." *Journal of Economic Surveys* 2(1): 174-205.

Comola, M. and L. Mello. (2013). "Salaried Employment and Earnings in Indonesia: New Evidence on the Selection Bias", *Applied Economics* 45: 2808-2816.

de Mel, S., D. McKenzie, & C. Woodruff (2008). "Who are the Microenterprise Owners?: Evidence from Sri Lanka", World Bank Working paper 4635.

Dubin, J. A. & McFadden, D. L. (1984). "An econometric analysis of residential electric appliance holdings and consumption.". *Econometrica, 52*(2), 345-362.

Duflo, E. (2001). “Schooling and Labor Market Consequences of School Construction in Indonesia : Evidence from an Unusual Policy Experiment”, *American Economic Review* 91:795–813.

Fields, G. (1980). “Education and Income Distribution in Developing Countries: A Review of the Literature.” World Bank working paper.

Fitzgerald, J., P. Gottschalk, & R. Moffitt. (1998). “An analysis of sample attrition in panel data.” *Journal of Human Resources, 33*(2), 251-299.

Foster, A. (2012). “Creating Good Employment Opportunities for the Rural Sector.” *Asian Development Review* 29(1), 1-28.

Haggblade, S., D. Mead and R. Meyer (2007). “An Overview of Programs for Promoting the Rural Nonfarm Economy,” in S.Haggblade, P.B.R. Hazell, and T. Reardon, eds., *Transforming the Rural Non-farm Economy* (Baltimore: Johns Hopkins University Press).

Hellerstein, Judith and Melinda Morrill. (2011). “Dads and daughters: The changing impact of fathers on women’s occupational choices”, *Journal of Human Resources* 46: 333-372.

Huffman, W.E. and P.F. Orazem. (2007). “Agriculture and Human Capital in Economic Growth: Farmers, Schooling and Nutrition”, Chapter 43 in: Evenson, R. and Pingali, P. (Eds.) , Handbook of Agricultural Economics Vol. 3., Elsevier.

Kijima, Y., Lanjouw, P. (2005). “Economic Diversification and Poverty in Rural India”, *Indian Journal of Labor Economics* 48(2).

Kramarz, F. and O.N. Skans. (2011). "With a Little Help from my Parents? Family Networks and Youth Labor Market Entry", Working paper, Center for Research in Economics and Statistics (CREST).

Lewis, A. (1954). "Economic Development with Unlimited Supplies of Labor", *The Manchester School* 22: 139-191.

Magruder, J.R. (2010). "Intergenerational Networks, Unemployment, and Persistent Inequality in South Africa", *American Economic Journal: Applied Economics* 2(1): 62-85.

Mead, D.C., Liedholm, C., (1998). The Dynamics of Micro and Small Enterprises in Developing Countries. *World Development* 26 (1), 61-74.

Mondragon-Velez, C. and X. Pena-Parga. (2008). "Business Ownership and Self-Employment in Developing Economies: The Colombian Case" in J.Lerner and A. Schoar, eds., *International Differences in Entrepreneurship* (Cambridge, MA: National Bureau for Economic Research): 89-127.

Oyelere, R. U. (2010). "Africa's education enigma? The Nigerian story", *Journal of Development Economics 91*(1): 128–139.

Reardon, T. (1997), "Using Evidence of Household Income Diversification to Inform Study of the Rural Non-farm Labor Market in Africa", *World Development, 25*(5), 735-747.

Schoar, A. (2010). "The Divide Between Subsistence and Transformational Entrepreneurship", *Innovation Policy and the Economy, 10*(1), 57-81.

Timmer, C.P. (1988). "The Agricultural Transformation." In H. Chenery and T.N. Srinivasan eds., *Handbook of Development Economics,* Vol. 1. Amsterdam: North-Holland, pp. 275-331.

Timmer, C.P. (2002). "Agriculture and Economic Growth." In B. Gardner and G. Rausser, eds., *Handbook of Agricultural Economics, Vol. IIA*. Amsterdam: North-Holland. 1487-1546.

Timmer, C.P. (2009). "A World without Agriculture: The Structural Transformation in Historical Perspective", Henry Wendt Lecture, American Enterprise Institute, 2008.

Wooldridge, J. M. (2002). *Economic Analysis of Cross-sectional and Panel Data*, Cambridge MA: MIT Press.

CHAPTER 2

FARM AND NON-FARM OCCUPATIONAL AND EARNING DYNAMICS IN RURAL THAILAND

2.1 Introduction

The rural non-farm economy (RNFE) is increasingly seen as a pathway out of poverty in low- and middle-income countries. As land becomes increasingly scarce, a transition to the rural non-farm sector becomes essential for many land-constrained rural households, a natural part of the 'agricultural transformation' intrinsic to economic development (Timmer 1988, 2002). Firms and activities in the RNFE provide essential linkages in the development process between agriculture and the macroeconomy, becoming a key contributor to increasing rural incomes, reducing poverty, and stimulating economic growth (Timmer 2002). Thus most rural households earn at least some income from non-farm sources, as non-farm workers, operating non-farm businesses, or both (Reardon 1997, Foster 2012).

The growing number of empirical studies related to the RNFE can be divided into two general groups. The first group investigates the determinants of RNFE participation, either at household or individual levels (Gibson and Olivia, 2010; Jonasson and Helfand, 2010). The second group focuses on the impacts of RNFE participation on household income, rural poverty, and inequality (Reardon et al., 2000; Ferreira and Lanjouw, 2001; Cherdchuchai and Otsuka, 2006; Matsumoto et al., 2006; Hung et al., 2010). The literature relies overwhelmingly, however, on (repeated) cross-sectional evidence. The dynamic role of the RNFE and occupational transitions on rural household earnings has yet to be investigated intensively; Block and Webb (2001), Bezu et al. (2012) and Bezu and Barrett (2012) are notable exceptions that use longitudinal household-level data from Ethiopia, while Foster (2012) explores related issues using household-level panel data from India. Past research suggests that those engaged in highly productive non-farm activities typically enjoy upward earnings mobility (Barrett, Reardon and Webb 2001; Block and Webb 2001; Lanjouw, 2001; Bezu et al., 2012; Bezu and Barrett, 2012). Of course, it is also likely that individuals with higher initial wealth and human capital are more able to engage in high-return non-farm activities and benefit most from the RNFE, so there could be significant selection effects involved in this oft-found association (Barrett et al., 2005). As Banerjee and Newman (1993) theorize, in the presence of capital market imperfections, the ex ante poor tend to choose wage labor while the ex ante rich become entrepreneurs. Banerjee and Newman also emphasize the interplay between 'the distribution of income and wealth' and 'the dynamics of occupational choice', suggesting that people in developing countries do not have free choice over their occupations but rather face significant structural constraints.

Evidence from many countries reveals considerable heterogeneity in the RNFE. But most household businesses consist either of self-employed enterprises without paid, non-family employees or small-sized firms with limited firm expansion (Fafchamps 1994; Haggblade et al., 2007). These businesses face several constraints, such as access to capital, skilled labor, entrepreneurial ability, and government registry requirements. Subsistence self-employment does not automatically transition into the enterprise growth that increases both the business owner's household income and employment within their region (Mondragon-Velez and Pena-Parga, 2008; de Mel et al., 2008; Schoar, 2010). It therefore seems important to differentiate between non-farm self-employment without hired workers and those household enterprises that hire non-family members, which we term entrepreneurs. Little is known empirically about the earnings transitions between farm work, rural non-farm employment, and rural non-farm self-employment – with or without employees, especially with adequate controls for prospective unobserved heterogeneity associated with selection into distinct occupational groups.

The same sort of differentiation exists and similar transitions are occurring within the farming sector in many developing countries. Landless rural households commonly must rely on unskilled employment on others' farms to earn a meager living. These farmworkers are commonly the poorest members of rural communities. Farmers who neither work for others nor employ paid non-family workers historically represent a large-share of semi-subsistence producers. These are typically the 'smallholders' around whom much of the rural and agricultural development discourse revolves. More skilled farmers with access to capital often expand their operations, adding paid employees as their transition from semi-subsistence to commercial cultivation. But the transitions among farm worker, self-employed farming and farm employer status, as well as among those non-farm categories and three analogous occupational categories in the RNFE, remain poorly documented.

This study helps to fill these gaps by exploring rural farm and non-farm occupational and earnings dynamics in rural Thailand, differentiating between employment, self-employment and entrepreneurship (i.e., an employer of paid non-family workers) in both the farm and non-farm sectors. More explicitly, the research questions this paper explores are as follows. First, what patterns of occupational transitions exist among farm workers, self-employed farmers, farm employers, non-farm employees, the non-farm self-employed, and non-farm employers in rural Thailand? Second, how do occupational transitions affect directional earnings mobility? Which occupational shifts – e.g., from farm to non-farm employee or from non-farm self-employed to non-farm employer – are associated with people increasing or decreasing earnings when controlling for other characteristics?

There have been a few previous studies on occupational mobility in developing countries (Fuwa, 1999; Quadrini, 2000; Mondragon-Velez and Pena-Parga, 2008). Mondragon-Velez and Pena-Parga (2008), in particular, explore the transitions between unemployed, wage-earner, self-employed and business owner status in seven main cities in Colombia. They mainly focus on the determinants of entry into and exit from urban self-employment and business ownership. They find that most newly self-employed and entrepreneurs transition from wage employment rather than from unemployment. However, they find extremely low transitions from self-employment to entrepreneurship (and vice versa). In studies of the determinants of income mobility, Cichello et al. (2005), Woolard and Klasen (2005), and Fields et al. (2005) found that the conditional effects of occupation and sector of employment were statistically significant in South Africa and Latin America.

This paper uses the nationally representative Thai Socio-Economic Survey (SES) panel data collected annually between 2005 and 2007, and the subsequent round in early 2010. This study thus uses more rounds of nationally representative panel data, with far more individual observations over a longer period, than any of the prior RNFE studies. This enables us to employ multiple empirical approaches, some of which would not be possible with simply two observations per individual or a much more modest number of observations, in order to more robustly identify the effects of occupational transitions on earnings dynamics in a nationally representative sample. It also enables us to differentiate among alternative farm and non-farm occupations in a way that matters fundamentally to rural development policy debates.

Little is known about earnings and occupational mobility in rural Thailand. Isvilanonda et al. (2000) indicate the growing importance of income from non-rice crops and non-farm activities, using survey data from six villages in 1987 and 1998. They find that while the number of the poor declined, income inequality has increased. Cherdchuchai and Otsuka (2006), using the same baseline survey data in 1987 and a new survey in 2004, investigate a structural shift of household income away from farm to non-farm income sources. Unlike Isvilanonda et al. (2000), they find that non-farm employment expansion reduces the income gap and the difference in poverty incidence between prosperous and poor regions. However, neither of these two papers investigates earnings mobility as it relates to occupational transitions.

We find significant occupational transitions over the course of just five years, mainly involving moves into farm self-employment and non-farm (salaried or wage) employee positions rather than into farm laborer, non-farm self-employment or farm or non-farm employer positions. A fairly strict ordering exists among different occupational groups, with farm workers' earnings distribution first-order stochastically dominated by that of self-employed farmers, which is itself dominated by the farm employer earnings distribution each year. No robust dominance ordering exists between farm employers and the non-farm self-employed, but the earnings distributions of non-farm employers and employees stochastically dominate those of the non-farm self-employed (without employees) and of all the farm sector earnings distributions. Given such an occupational ladder, transitions from farming into non-farm employment therefore result in statistically significant income gains, on average, while moves into farming are associated with reduced earnings.

That core finding of an occupational ladder is reinforced by directional earnings mobility regression analysis when tracking the same individuals over time. Only a small number of individuals become non-farm employers, the most remunerative occupation group, reflecting the difficulty inherent to establishing and maintaining a business with employees. Moreover, less than one percent of these household enterprises employ ten or more family members (Chawanote 2013), indicating limited employment generation potential through household-based non-farm enterprises in rural Thailand. Our findings suggest that promoting rural non-farm employment by attracting established businesses, government or not-for-profit agencies may be more important to rural poverty reduction than promoting rural non-farm self-employment in the hope that this leads to entrepreneurial rural non-farm job creation and higher rural household incomes.

2.2 Data and background

Thailand is a lower middle-income country by the World Bank's classification, with GDP per capita of $8,004 in 2009. According to the World Bank (http://data.worldbank.org/country/thailand), the $2/day per person poverty headcount ratio was 11.5 percent of population in 2004, down from 16.7 percent after the 1997-98 financial crisis. The labor force participation rate was 73.2 percent of the total population ages 15 and above. Roughly 1.3 percent of the total labor force reported being unemployed between 2005 and 2009. Approximately 67 percent of the population from 2005 to 2009 lived in rural areas, with a steadily declining share employed in agriculture.

The Thai SES panel data were collected by the National Statistical Office (NSO) of Thailand in 2005 – 2007 and 2010. The repeated cross-sectional rounds of the well-respected SES have been used frequently by leading researchers (e.g., Schultz 1990, Paxson 1992, Mammen and Paxson 2000, Giné and Townsend 2004, Felkner and Townsend 2011). Beginning in 2005, NSO began tracking households and split-off individuals from sample households to create proper panel data, although these panel data appear not to have been exploited much, if at all.[1] We therefore take particular care to explore attrition patterns and their implications (Appendix A), so as to enhance the usefulness of this rich longitudinal data set for future researchers.

For the first three rounds (2005-7), the survey was recorded in May, while the last (2010) round was surveyed in January. The survey has two main segments: i) household information on every member in the household, and ii) individual information on household members aged 15 years or older. Part one includes general

[1] The SES panel data have been used for internal government reports and by some research institutes in Thailand. But we can find no English language publications that exploit these important data, likely due to the facts that printed information on the panel is only available in the Thai language, thereby sharply limiting awareness of their availability, and that the data are not freely available to the public, but require a purchase contract with NSO approving their release and use.

information on household members, household characteristics and assets, and income from agriculture. Part two includes survey questions on education, health care, employment, incomes, expenditures, financial status (debt and savings), migration, and opinions on public policies. The survey covers every province in Thailand and randomly selects blocks of districts, sub-districts and villages, and finally selects ten households per village as in a two-stage stratified random sampling.[2] All statistics we report are adjusted for sampling weights.

Table 2.1 summarizes the Thai SES panel data.[3] The 2005 round surveyed 6,000 households with a total of 16,310 individuals and 9,897 individuals in rural areas. All individuals age 15 and over were tracked in the following years' surveys. Any adult who left the core household was tracked so long as they remained within Thailand and a new address for that split-off individual could be found by the survey team. Some individuals are missing from one round, but reappear in later rounds once they could be tracked again. We use only the balanced panel, in other words, only individuals present in all four rounds of the SES. Due to split-off households and attrition, the total number of individuals aged 15 years or older surveyed in all four rounds is 12,758, of whom 7,831 lived in rural areas. Given the substantial attrition, we take special care to control for the possible bias this might introduce (Appendix A).

Since the Thai SES panel only surveys at household and individual level, we match it with another village-level dataset. A rural development census, the National Rural Development (NRD) data set, was collected at the village level by the

---

[2] Each rural sub-district in the SES panel data has only one village.
[3] Further explanation of the rural sample across years and attrition issues are provided in Appendix 2A.

Community Development Department of Thailand.[4] NRD data that match the Thai SES panel data are only available for April to June 2005 and April to May 2007 and 2009. The data cover general conditions of the village and local economy, including the availability of public services and infrastructure, health and sanitation, village educational achievement, and agroecological conditions.

Table 2.1: Summary information on SES sample

Year	Individuals	Rural Individuals	Households	Rural Households
2005	16,310	9,897	6,000	3,680
2006	16,542	10,208	6,020	3,752
2007	16,490	10,350	5,955	3,783
2010	17,045	10,915	6,244	4,002
Observed all 4 years	12,758	7,831	5,229*	3,362*
Observed all 4 years (age $\leq$ 70)	11,484	7,000		
% Rural by 3 year total		61.4%		64.3%

* Based on the household ID that was recorded in 2005.

2.2.1 Definition of rural non-farm employment

We use only the Thai SES panel data on individuals who were employed in rural areas, including unpaid workers for household businesses, and those who were 15-70 years old. Over the five-year period of SES data collection, the unemployment rates in the rural areas included in the study ranged from 0.5 to 1.1 percent while employment rates ranged from 76.1 to 79.9 percent.[5] In the employment section of the SES, respondents were asked for their primary occupation, work status, and company

[4] The data are distributed by the University of Chicago-UTCC Research Center, Bangkok, Thailand.

[5] Other categories in the survey are waiting for seasonal work, looking for work, retired, long term illness and disabilities, caring for other household members, and going to school. We only focus on employment status since this better represents a group of earners and eliminates the possible variation in occupational transitions that would come with seasonal work.

size[6] for each of up to three jobs that they had worked in the past 12 months. The first job recorded in the dataset reflects the individual's current main job at the time of survey.[7] It should not be affected by seasonality for those who are in the farming sector, given the survey timing. May is the beginning of the rice cultivation season in Northern and Central regions, rice harvesting season in Southern region, and other harvesting season for fruits in Eastern Thailand. Even though the 2010 survey round was in January, it is a main cultivation period for tapioca/cassava, cane, and other similar crops. The options for primary occupation in the survey are farmer/fisherman (crops, livestock, aquaculture, fishery, hunting and gathering), production (handicrafts and basic manufacturing), production (industry), merchandise/own business, government/state enterprise employee, company/business employee, and general worker/laborer. The work status question includes options for employer, self-employed without employees, working without pay for household business, government employee, state enterprise employee, private company employee, and cooperative group. These two questions – primary occupation and work status – are used to separate non-farm activities from farm activities at the individual level and to differentiate among workers, the self-employed and employers in the farm and non-farm sectors.

---

[6] The company size, measured as the total number of workers including the owner, is categorized as 1 worker (i.e., no employees), 2-9 workers, 10-50 workers, 51-100 workers, 101-200 workers, 201-500 workers, or over 500 workers.

[7] Five percent of 1st jobs were reported as having ended in the 12 months prior to the survey while around 99 percent of the 2nd and 3rd jobs were reported as having ended in the previous year. The length of time in 2nd jobs was consistently an order of magnitude shorter than in 1st jobs and time spent in 3rd jobs was only 16-25 % that in 2nd jobs. The 2nd and 3rd jobs reflect seasonal jobs or jobs that ended before the current, primary one. Dropping these short-term, temporary positions makes no qualitative difference to the analysis we report here.

Unfortunately, the household survey does not directly identify a respondent's employer, so we cannot match employees with employers. We do, however, know the size distribution of individual respondents' employers. As reflected in Table 2.2 for 2005 and 2010 (the 2006 and 2007 data exhibit qualitatively identical patterns), at most one-third of rural non-farm employees work for private businesses with fewer than ten employees (grey-shaded cells).[8] Consistent with evidence from high-income countries (e.g., Hurst and Pugsley 2011), parallel analysis of household enterprise data from the SES panel (Chawanote 2013) finds that less than one percent of household-owned enterprises in rural Thailand employ ten or more workers and very few of these enterprises exhibit any statistically significant employment growth over the 2005-10 period.[9] The striking mismatch between the jobs created by rural household enterprises and the jobs held by those who work for a salary or wages in the rural non-farm economy carries important policy implications. Donors' and governments' present emphasis on promoting rural household non-farm entrepreneurial activity might not offer an adequately broad platform to facilitate agrarian transformation and rural earnings growth.

[8] Furthermore, employees' reports of small-sized state owned enterprise or government employers almost certainly refer to respondents' immediate department/unit rather than to the entire agency. If similar underreporting of private firm size occurs, that reinforces our finding.
[9] Thailand's most recent (2007) establishment enterprise survey, which covers only the manufacturing sector, reports that only 5.7 percent of enterprises employ more than 15 workers.

Table 2.2: Sector of individual non-farm employment, by employer size (percent)

2005	Government	State owned enterprise	Private sector	Total
2-9 workers	3.6	0.1	36.0	**39.8**
10-50 workers	8.4	0.4	19.1	**27.9**
51-100 workers	2.1	0.0	3.5	**5.7**
101-200 workers	1.4	0.2	3.8	**5.4**
201-500 workers	0.6	0.0	4.8	**5.4**
over 500 workers	6.6	0.8	8.5	**15.8**
Total (n=1,540)	**22.7**	**1.5**	**75.8**	**100.0**
2010				
2-9 workers	3.6	0.2	31.4	**35.2**
10-50 workers	12.3	0.6	19.5	**32.4**
51-100 workers	3.0	0.2	6.0	**9.2**
101-200 workers	1.5	0.3	4.9	**6.7**
201-500 workers	0.6	0.1	5.7	**6.4**
over 500 workers	3.7	0.2	6.3	**10.1**
Total (n=1,355)	**24.7**	**1.6**	**73.7**	**100.0**

Each cell reports a percentage of total non-farm employment.

The rural non-farm sector includes all economic activities in rural areas except primary production in agriculture, livestock, fishing and hunting, and thus includes any employment in manufacturing, mining, trade, construction, transportation, communications, government and services (Lanjouw, 2001; Haggblade et al., 2002). Using this definition, those who reported their primary occupation as being anything other than farmer/fisherman are considered as working in the non-farm sector. Conversely, only those who reported their primary occupation as farmer/fisherman are considered as working in the farm sector.

Previous studies that decompose rural non-farm employment have categorized it as either low-productivity wage labor or high-productivity salaried work or self-employment (e.g., Barrett et al., 2005; Jonasson and Helfand, 2009; Bezu et. al., 2012). In this study, however, entrepreneurship status is classified separately from self-employment without employees as this differentiates what are sometimes referred to as "subsistence" from "transformational" entrepreneurs, with the latter being the prospective source of new RNFE jobs (Schoar, 2010). Any respondent who employed non-family members in any (farm) non-farm activity is considered a ('farm employer') 'non-farm employer' or 'entrepreneur', while anyone self-employed without employees, working without pay for a household business, or working in a cooperative group is grouped into the 'self-employment' category, either self-employed farming for those with primary occupation in farming/fishing or non-farm self-employment otherwise. Both employers and the self-employed refer to those who operate their own business and receive business profits as their primary earnings. Finally, 'employee' includes salaried and waged workers, i.e., those employed by the government, state enterprises, or private companies or not-for-profit agencies, and who have no claim to business profits. Farm workers are employees working for a farmer; non-farm employees work for an enterprise outside the farm sector.

Table 2.3 summarizes individuals' work status in rural Thailand. The percentage of workers in each occupation changed only slightly between 2005 and 2010. Although farmers and farm workers represent a plurality of rural Thai workers and self-employed farmers are the single largest category in 2007 and 2010 and over the four year panel cumulatively, more people are employed primarily in non-farm occupations. Non-farm employees account for the largest proportion of non-farm sector workers, while non-farm employers account for only one percent of the total employed population in rural Thailand. Similarly, farm employers account for only eight percent of the total farming population (farmers plus self-employed farmers) on average over the 2005-10 period. More than 90 percent of farm and non-farm business owners in rural Thailand do not create jobs outside the entrepreneur's household and of those who become employers, less than one percent create 10 or more jobs. This is an important point largely missed in the literature and in contemporary policy dialogues, which emphasize promoting entrepreneurship to ignite employment in the rural non-farm economy. That supposed engine of growth seems to have relatively little power.

Table 2.3: Work status in rural areas

Work Status (ages 15-70 years)	2005		2006		2007		2010	
	N	%	N	%	N	%	N	%
Unemployed	72	1.0	75	1.1	70	1.0	38	0.5
Other work status*	1,600	22.9	1,474	21.1	1,335	19.1	1,379	19.7
Employed status in each year	5,328	76.1	5,451	77.9	5,595	79.9	5,583	79.8
- Employed < 4 waves	1,238	17.7	1,361	19.4	1,505	21.5	1,493	21.3
- Employed status in all 4 waves	4,090	58.4	4,090	58.4	4,090	58.4	4,090	58.4
• Self-employed farmer	1,415	20.2	1,514	21.6	1,623	23.2	1,734	24.8
• Farm worker	122	1.7	66	0.9	27	0.4	58	0.8
• Farm employer	225	3.2	171	2.4	104	1.5	127	1.8
• NF self-employed	733	10.5	727	10.4	781	11.2	738	10.5
• NF employee	1,540	22.0	1,546	22.1	1,496	21.4	1,362	19.5
• NF employer	53	1.3	64	0.9	56	0.8	65	0.9
• Missing data on occupation	2	0.05	2	0.03	1	0.01	3	0.04
Total	7,000		7,000		7,000		7,000	

*Other work status includes waiting for seasonal work, unemployed, looking for work, retired, long term illness/disability, caring for other household members, student, and others.

In the analysis that follows, we focus on the earnings and occupational dynamics of only those rural working age adults (15-70 years old) who were employed and surveyed in all four SES rounds, so as to avoid conflating transitions between unemployment and employment with transitions among occupations. This introduces the possibility of attrition bias, either due to exits from the sample – due to outmigration, death, unavailability, or another reason – or because of one or more periods of unemployment during the SES rounds. Appendix 2A explores the possibility of non-random attrition in detail, demonstrating that attrition indeed appears non-random, although the attrition-corrected regression results reported in the main body of the paper are not statistically significantly different from the uncorrected results in Appendix 2A, Table A5-A7.

2.2.2 Earnings

Individual earnings are decomposed by source: farm earnings, non-farm business profits, and wages or salaries. Farm earnings and non-farm business profits are recorded at the household level. We use individual work hours per week in each enterprise to assign individual farm income to individual household members based on their share of total family labor time allocated to the farm enterprise. Similarly, non-farm business profits are allocated to all self-employed members in the household proportional to time self-employed members work in the household non-farm enterprise. Wage and salary earnings are already recorded at the individual worker level, where we also know the sector of employment. All earnings are adjusted for the consumer price index for each region of Thailand to put them in real 2007 baht terms.[10]

---

[10] Consumer price index data by region are reported by Thailand's Ministry of Commerce (www.moc.go.th).

We focus on structural occupational transitions and individual earnings mobility 2005-2007-2010. This allows us to use village-level controls available from the NRD (which, as indicated previously, was not fielded in 2006 and 2008). Plus, the longer spell length in the dynamics analysis minimizes the role of transitory shocks and measurement error, reducing the possibility of overstating structural economic mobility (Naschold and Barrett 2011).

Table 2.4 shows mean earnings by quartiles conditional on each occupation and with the lowest and highest one percent of earnings in each year cut off so as to eliminate extreme outliers likely to reflect measurement error.[11] On average, non-farm employers enjoy the highest earnings while farm workers receive the lowest earnings. Non-farm employees earn more on average than those engaged in non-farm self-employment in every quartile, and farm employers in the fourth quartile earn more on average than do individuals with non-farm self-employment. Compared against the 2007 rural poverty line of 1,333 Baht[12] per capita per month, within the first quartile only non-farm employees and employers on average earn above the poverty line. But even the second quartile of self-employed farmers and, in most years, even the third quartile of farm workers, having mean earnings below the poverty line. Twenty percent of the rural employed fall under the poverty line; almost eighty percent of the rural poor are in farming. However, this is based solely on occupational earnings, excluding income from other sources such as remittances, incomes from house/land

[11] Note that the core qualitative results in the analysis that follow are robust to inclusion of these tail observations, but parameter estimates become considerably less precise when we incorporate these suspicious looking values. So we favor the analysis based on the trimmed sample.

[12] In 2007, there were 33.72 baht per US dollar, referencing from the Office of the National Economic and Social Development Board (NESDB).

Table 2.4: Mean individual earnings (per month) by quartile and occupation

2005	1st quartile	2nd quartile	3rd quartile	4th quartile	Overall
Farm self-employed	437	1,306	2,554	7,289	3,385
	(242)	(240)	(465)	(4,293)	(3,679)
Farm worker	0	79	533	2,116	645
	(0)	(44)	(263)	(1,530)	(1,150)
Farm employer	676	1,995	3,716	12,083	5,219
	(347)	(445)	(809)	(6,253)	(5,792)
NF self-employed	1,051	2,741	4,866	12,203	5,503
	(649)	(439)	(822)	(6,618)	(5,484)
NF employee	2,220	4,188	6,167	15,827	7,716
	(800)	(521)	(760)	(7,163)	(6,667)
NF employer	3,295	6,450	9,554	22,287	10,471
	(1,379)	(752)	(1,383)	(6,817)	(8,197)

2007	1st quartile	2nd quartile	3rd quartile	4th quartile	Overall
Farm self-employed	389	1,147	2,358	7,020	3,250
	(223)	(251)	(509)	(4,155)	(3,619)
Farm worker	-34	NA	NA	2,554	446
	(64)	NA	NA	(2,020)	(1,326)
Farm employer	887	2,582	5,865	14,253	6,215
	(615)	(611)	(1,420)	(7,071)	(6,352)
NF self-employed	1,060	3,073	5,837	12,868	6,178
	(648)	(637)	(982)	(5,812)	(5,484)
NF employee	2,423	4,550	6,688	16,907	8,542
	(879)	(482)	(901)	(8,277)	(7,461)
NF employer	4,305	9,674	16,601	24,278	13,286
	(2,202)	(1,508)	(2,301)	(6,130)	(8,033)

2010	1st quartile	2nd quartile	3rd quartile	4th quartile	Overall
Farm self-employed	331	1,390	2,831	9,090	3,867
	(450)	(301)	(647)	(5,966)	(4,809)
Farm worker	-18	NA	196	2,758	975
	(74)	NA	(3)	(2,523)	(1,989)
Farm employer	1,211	3,086	5,463	19,660	7,999
	(552)	(491)	(1,200)	(12,029)	(9,740)
NF self-employed	940	3,001	5,676	13,348	5,808
	(593)	(688)	(863)	(5,484)	(5,528)
NF employee	2,839	5,309	8,383	21,936	10,035
	(1,123)	(628)	(1,377)	(9,124)	(9,045)
NF employer	3,656	6,537	14,038	25,549	11,407
	(806)	(1,127)	(3,295)	(7,971)	(8,934)

Units are 2007 Thai baht. The lowest and highest one percent of the rural income percentiles in each year have been omitted. Quartiles are based on employed status in each year. Standard deviation is shown in parentheses for each quartile by occupation. Estimates not available (NA) because there are no observations in that quartile.

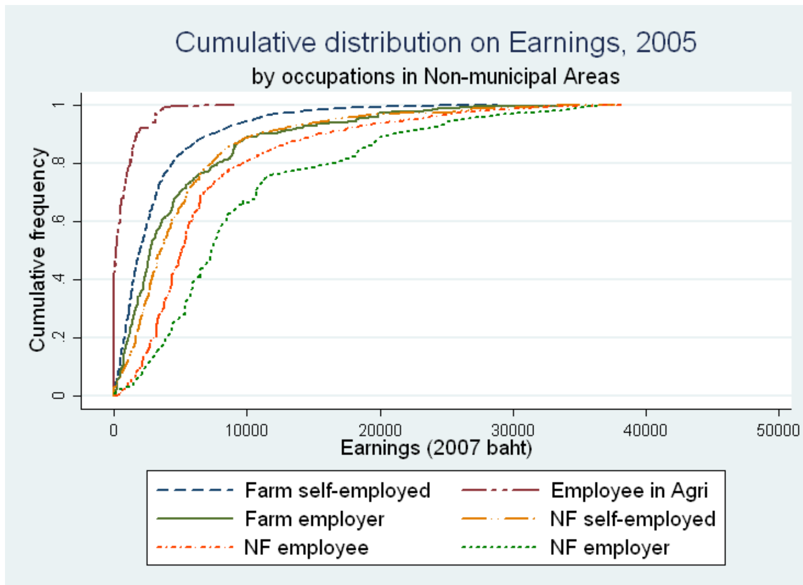
lending, or returns from financial assets. The poor seemed to be affected most by the country's 2008-9 economic downturn as the earnings averages in the first quartile in 2010 dropped from 2007, whereas the highest quartiles still enjoyed an increase in earnings on average. The economic slowdown also had an impact on non-farm businesses since earnings of both non-farm self-employed and employers in 2010 fell slightly from 2007.

Figure 2.1 presents the cumulative frequency distributions of earnings by occupational status and sector in 2005 and 2010. Every other category's earnings distribution first order stochastically dominates that of farm workers in each period, and the farm employer and each non-farm category's earnings distribution first order stochastically dominates that of self-employed farming in each period.[13] These are striking results that clearly underscore the relative undesirability of farm work and semi-subsistence agriculture.

Perhaps even more striking, however, no stochastic dominance ordering appears between the earnings distributions of farm employers – the largest, most commercial farm operations in the country – and the non-farm self-employed in any survey year. But each is at least second-order stochastically dominated by both the non-farm employee and employer earnings distributions in each year and first order dominate them in several years.These orderings are consistent with the results presented in Table 2.4, where the most desirable remunerative is non-farm employer followed by non-farm employee, non-farm self-employment, farm employer, self-employed farming, and farm worker last of all, in that order. There is no statistically

[13] We use Davidson and Duclos' (2000) test to confirm all the stochastic dominance results reported here.

significant stochastic dominance ordering between the two dominant (non-farm employee and non-farm employer) distributions due to a small number of low earnings draws among non-farm employers. But mean earnings for non-farm employers are considerably higher, albeit with the gap closing over the 2005-10 period.



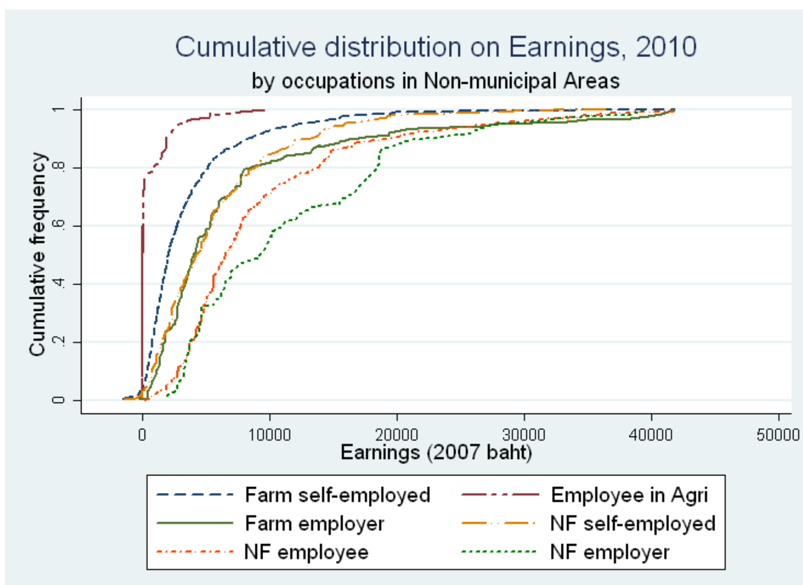


Figure 2.1: Cumulative distribution by occupation, 2005 and 2010

2.3 Earnings changes and occupational transitions

Having already observed a clear earnings distribution ordering among occupations in each year, we would expect that any transitions from farm work or from self-employed farming into rural non-farm occupations should be associated with increased earnings, as should transitions from farm employer or non-farm self-employment to non-farm employee or non-farm employer status. Conversely, transitions into farming, or into non-farm self-employment from the other two non-farm occupational categories, should be associated with reduced earnings, as should any transitions into being a farm worker or transitions into self-employed farming from any occupation other than farm worker. This intuition is confirmed by extending the repeated cross-sectional analysis to intertemporal transitions.

The transition matrices presented in Table 2.5 describe movement across farm and non-farm employment categories. The percentage change is calculated to show how occupational status in 2005 (row) changed by 2010 (column). Other than for farm and non-farm employers and farm workers, work status primarily remains the same across the five years, with 63-74 percent of each group remaining in their original occupational sector. But almost 30 percent of 2005 non-farm employers had shed their employees and converted to merely self-employed status by 2010 while around 40 percent maintained their non-farm employer status for those five years, although the sub-sample size is small. Farm employers have an even tougher time maintaining jobs, as only 19 percent of those who had paid workers in 2005 still employed non-family members in 2010. Nearly 70 percent of those who were farm employers in 2005 had reverted back to self-employed farming by 2010. Although these larger, more commercially oriented enterprises are the most remunerative within their respective sectors, it is clearly difficult to maintain a household business that employs others, whether in the farm or non-farm sectors.

Table 2.5: Transition matrix of occupational changes 2005/2010

Farm and Non-farm	Farm employment 2010			Non-farm (NF) employment 2010			
employment 2005	Self-employed	Employee	Employer	Self-employed	Employee	Employer	Total
Farm self-employed							
- *Number*	1,040	18	47	117	173	7	1,412
- *Row percentage*	73.7	1.3	4.0	8.3	12.3	0.5	100.0
Farm employee							
- *Number*	62	6	2	6	46	0	122
- *Row percentage*	50.8	4.9	1.6	4.9	37.7	0.0	100.0
Farm employer							
- *Number*	154	1	43	11	14	2	225
- *Row percentage*	68.4	0.4	19.1	4.9	6.2	0.9	100.0
NF self-employed							
- *Number*	139	3	11	462	93	24	732
- *Row percentage*	19.0	0.4	1.5	63.1	12.7	3.3	100.0
NF employee							
- *Number*	329	30	13	127	1,028	11	1,538
- *Row percentage*	21.4	2.0	0.9	8.3	66.8	0.7	100.0
NF employer							
- *Number*	8	0	1	15	8	21	53
- *Row percentage*	15.1	0.0	1.9	28.3	15.1	39.6	100.0
Total	1,732	58	127	738	1,362	65	4,082
- *Row percentage*	42.4	1.4	3.2	18.1	33.4	1.6	100.0

The total number of individuals here differs from that reported in Table 2 because some employed individuals are missing data on occupation

At the other of the earnings spectrum, farm workers are the individuals most likely to transition out of that status. Only five percent of those whose primary employment was as a hired farm worker in 2005 were still relying primarily on agricultural wage labor in 2010. More than 55 percent had escaped to self-employment, mainly in farming but also in the non-farm sector to a substantial degree. The extraordinarily high rate of occupational mobility out of hired farm labor underscores the unattractive earnings prospects of those forced to rely on agricultural wage labor for their primary livelihood.

As one would expect, transitions are more from farming into more remunerative non-farm employment. However, far more people slip from non-farm self-employment into farming than graduate into the most remunerative non-farm employee or employer positions. Likewise, almost 12 times more non-farm employees slip back into non-farm self-employment than graduate into becoming non-farm employers. The non-farm self-employed are more likely to transition into employer status than are those who did not previously run a non-farm business. But as in Mondragon-Velez and Pena-Parga (2008), we find an extremely low transition rate into being a non-farm employer, just 3.3 percent of the non-farm self-employed and less than one percent of each of the other four occupational categories.

These results strongly suggest a ‘gravity effect’ on the occupation ladder: it is easier to move down into lower-return occupations than up into higher-return ones, although it is relatively easy to take a single step up the ladder, from agricultural wage labor to self-employed farming. These transition patterns indicate the difficulty of starting and expanding a non-farm business or even securing paid non-farm employment, since the earnings distributions for those two occupational groups first order stochastically dominate the earnings distributions of the other two categories. It is even difficult for farmers to start or expand agricultural operations to employ non-family members or even to maintain a farm payroll. Constraints may include differences in physical and human asset endowments, access to finance, social connections, etc. We discuss these issues more in section 4 when we investigate the determinants of occupational transitions.

Table 2.6 presents the median, mean and standard deviation percentage real earnings changes associated with each transition. None of the earnings changes are statistically significantly different from zero, reflecting the considerable dispersion observed in unconditional earnings transitions. The mean and median patterns are similar in their directional changes. Individuals who remained in their initial occupational categories, except farm workers, enjoyed positive mean changes in earnings. Movement from farming into any non-farm employment generates earnings gains, on average, while moving into farm employment or self-employed farming is associated with earnings losses, on average. But note that of the roughly 40 percent of non-farm employers who maintain their business and employees over the course of five years, most suffered a decline in earnings over the 2005-10 period. This underscores the considerable challenge of maintaining, much less growing employment through nonfarm household enterprises in rural Thailand.

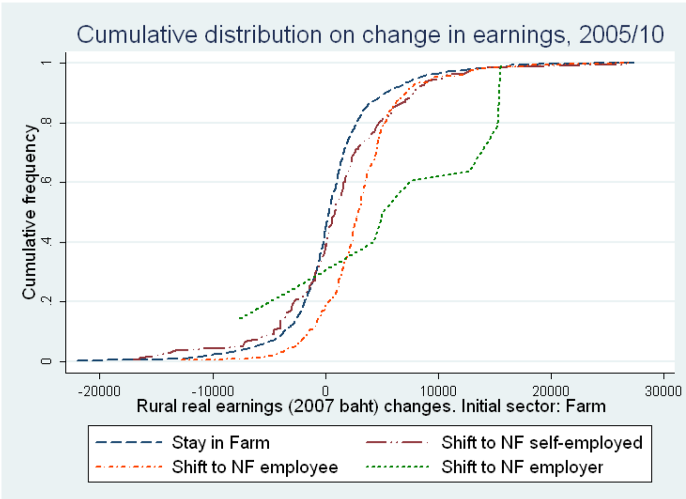
Table 2.6: Transition matrix of median and mean percentage change in earnings 2005/2010

Farm and Non-farm employment 2005	Farm employment 2010 Self-employed	Farm employment 2010 Employee	Farm employment 2010 Employer	Non-farm (NF) employment 2010 Self-employed	Non-farm (NF) employment 2010 Employee	Non-farm (NF) employment 2010 Employer
Farm self-employed						
- *Median of* ΔY	225	-782	2,164	887	2,754	7,499
- *Mean of* $\%\Delta Y$	1.11	-0.06	1.51	2.41	3.52	5.43
- *(s.d.)*	(3.08)	(1.55)	(2.90)	(5.21)	(5.56)	(7.79)
Farm employee						
- *Median of* ΔY	1,839	0	1,111	5,889	4,628	NA
- *Mean of* $\%\Delta Y$	3.84	-1.12	14.18	3.22	8.08	NA
- *(s.d.)*	(5.80)	NA	NA	(6.47)	(6.25)	NA
Farm employer						
- *Median of* ΔY	542	-2,579	1,098	4,428	2,222	15,261
- *Mean of* $\%\Delta Y$	1.03	-1.00	0.89	3.37	4.19	5.53
- *(s.d.)*	(2.84)	NA	(2.12)	(3.62)	(7.72)	(7.86)
NF self-employed						
- *Median of* ΔY	-479	-838	2,153	479	1,538	-190
- *Mean of* $\%\Delta Y$	0.82	-0.23	2.47	1.09	0.98	0.69
- *(s.d.)*	(3.14)	(0.65)	(4.33)	(3.29)	(1.97)	(2.61)
NF employee						
- *Median of* ΔY	-1,251	-4,329	1,931	-1,056	957	5,604
- *Mean of* $\%\Delta Y$	0.11	-0.74	1.60	0.36	0.32	1.02
- *(s.d.)*	(1.93)	(0.49)	(3.35)	(1.87)	(0.95)	(1.16)
NF employer						
- *Median of* ΔY	-2,209	NA	-1,306	416	2,279	-2,107
- *Mean of* $\%\Delta Y$	-0.27	NA	-0.34	-0.10	0.19	0.60
- *(s.d.)*	(0.84)	NA	NA	(0.60)	(0.78)	(1.30)

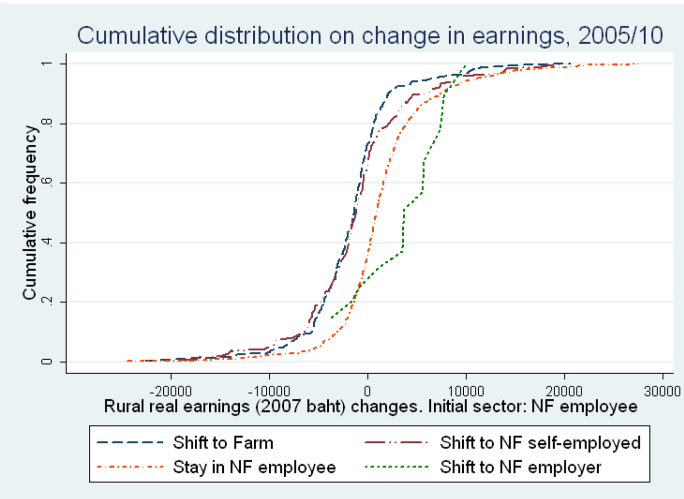
Earnings changes (ΔY) are in real Thai baht, adjusted by CPI for each region of Thailand with base year 2007. The reported statistics omit the top and the bottom one percent of the sample and correct for attrition weights as described in Appendix 2A.

We can generalize this analysis to explore the full distribution of earnings changes associated with each transition (Figure 2.2), dropping the lowest and highest one percent of earnings changes. From this point on, we begin aggregating the three farm sector occupations (agricultural wage laborer, self-employer farmer, and farm employer) because the small number of observations of farm employees and employers makes disaggregated, conditional mobility analysis infeasible for those subgroups. The fact that the farm sector occupations are strictly orderable internally and uniformly stochastically dominated by non-farm employment and non-farm entrepreneurship enables us to use this aggregation without losing important nuance that might matter for policy-related inferences.

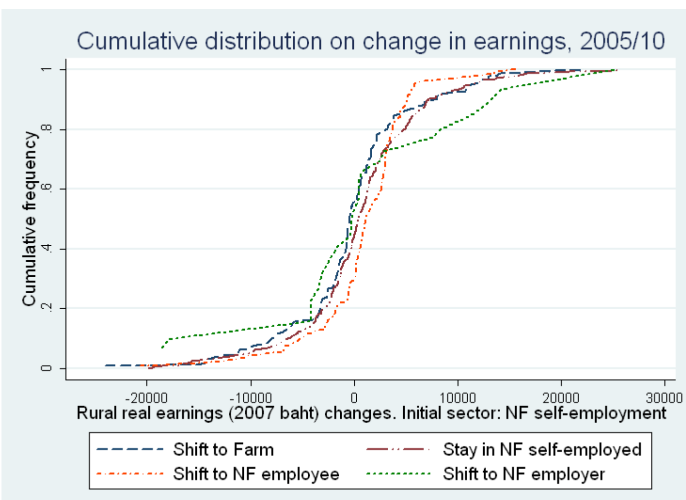
The plots in Figure 2.2 show that some earnings changes distributions first order stochastically dominate others, although there is no clear and consistent ranking among the distributions of earnings changes of initial employer positions based on stochastic dominance tests. For those initially in the farm sector, the transition to non-farm employee status first order stochastically dominates staying in the farm sector. However, none of these earnings changes distributions reveal statistically significant, second order stochastically dominant transitions into non-farm self-employment, nor consistently significant transitions into non-farm employee or employer status.



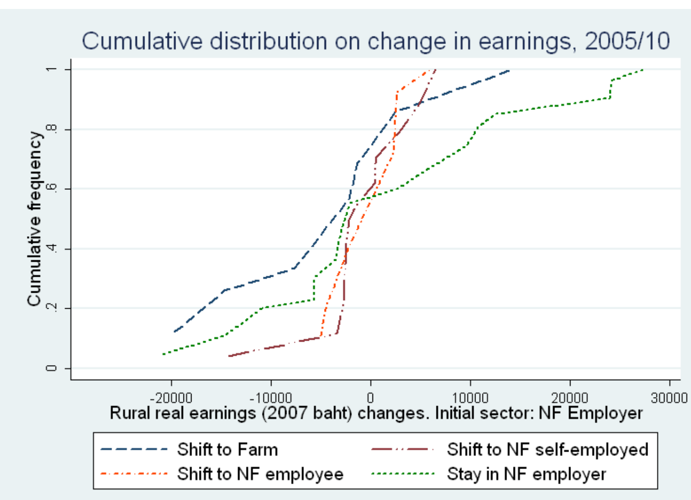
(a) Transitions from farm



(b) Transitions from non-farm employee



(c) Transitions from non-farm self-employed



(d) Transitions from non-farm employer

Figure 2.2: Cumulative distribution of change in earnings
by occupational transition between 2005 and 2007

2.4 Multivariate analysis of occupational shifts and earning mobility

2.4.1 Empirical model

Especially given the absence of an explicit earnings change ordering among occupational transitions and the non-random nature of those transitions, multivariate regression analysis can help us better understand how changes in earnings associate with farm and non-farm occupational shifts. We emphasize that in these observational data, it is exceedingly difficult to control for all prospective sources of unobserved heterogeneity that might generate selection effects or spurious correlation between occupational transitions and earnings dynamics. We can convincingly establish associations only. But by employing a range of controls and estimation techniques, each aimed at addressing a different source of prospective bias, we can check if the core qualitative results are robust to a range of statistical corrections that are each incomplete and imperfect but as a set offer a reasonably comprehensive approach to check the core results. The robustness of the findings and the quality of the data give us confidence that the strong and consistent statistical associations we find likely indicate a true causal relationship between occupational transitions and earnings dynamics in rural Thailand.

We employ a conditional mobility model in which change in earnings or change in log earnings are regressed on time-invariant and time-varying individual characteristics. In this class of model, changes in earnings are explained by initial earnings, gender, age, educational attainment, sector of employment, and geographic region, with occupation and sector of employment typically considered time-varying variables (Cichello et al., 2005; Fields, 2007). This framework allows us to explore how occupational shifts change earnings when controlling for other observable characteristics that are almost surely correlated with both earnings dynamics and occupational patterns. Following Fields (2007), the conditional micro mobility model is defined as:

$$\Delta \ln y_{it} = \alpha + \beta_1 \ln y_{i,t-1} + t \ln y_{i,t-1}\, \beta_2 + \Delta X_{it}\beta_3 + Z_i\beta_4 + \phi_i + \lambda_t + \varepsilon_{it} \qquad (2.1)$$

where $\Delta \ln y_{it}$ is the change in log reported real earnings from year t-1 to year t and ln $y_{i,t-1}$ is the prior survey year's log reported real earnings, included as a control for autocorrelation.[14] Because the periodicity of the SES panel changed, from one year revisits between the 2005, 2006 and 2007 rounds, to a three year revisit in the 2010 round, we do not impose a single autocorrelation parameter. Instead, we add interaction terms between the base year earnings and year dummies for the 2006-7 and 2007-10 transitions. Z_i denotes a matrix of time-invariant individual and household characteristics, as observed in the initial year. Both age and age squared are included to control for life cycle effects that should be reflected in a positive (negative) sign on the linear (quadratic) term. Education is recorded as the highest level completed, with dummy variables for primary school, secondary school, high school/vocational school, and college degree and above, with less than primary school or none as a base level. Gender is described with a dummy variable taking value one for females, and marital status is described with a dummy taking value one for married persons. Since the observations are at the individual level, a dummy for household head is also included, as well as family size. An initial year asset index and household owned agricultural

[14] We use a logarithmic specification because it substantially improves goodness of fit relative to using earnings levels.

land separately from the asset index are also included to control for household capital endowments.[15] ΔX_{it} denotes employment transition experiences, which are represented by dummy variables for fifteen possible transitions, with staying in farm work as a base case. Finally, ϕ is a vector of sub-district fixed effects, λ is a vector of time fixed effects, and ε_{it} is a mean zero i.i.d error term, corrected for clustering and potential heteroskedasticity.

We hypothesize that the sectoral transitions' coefficient estimates in the log earnings equation follow the same ordering found in the unconditional analyses reported in section 2.3, even after controlling for individual and household characteristics. Moreover, we can also test the differences between occupational transitions' coefficients, given the initial or previous job, for earnings changes associated with those occupational shifts, similar to testing for stochastic dominance in Figure 2. That is, transitions into (out of) farming, or into (out of) non-farm self-employment from the other two non-farm occupational categories should be associated with reduced (increased) earnings.

2.4.2 Empirical results

Table 2.7 provides descriptive statistics of these variables for the whole sample and for each group. Given each group in 2005, the mean of the asset index is the highest for non-farm employers and lowest for farmers, although there is not much

[15] The estimation details of the asset index, constructed using factor analysis following Sahn and Stifel (2003), are reported in Appendix 2B. The index includes number of rooms, housing materials, electricity, cooking fuels, water supply, toilet, number of durable goods (e.g., microwave, refrigerator, air conditioner, fan, television, radio, VCD-DVD player, washing machine, cable television, cell phone, landline, computer and internet), number of vehicles (motorcycles, cars, trucks, tractors), and livestock.

difference in means of the asset index between non-farm self-employment and non-farm employees. Non-farm employees have the highest proportion of college graduates as opposed to farmers that have the highest proportion of primary school graduates.

The estimation results, using 2005-6, 2006-7, and 2007-10 transitions and annualized log earnings changes, are reported in Table 2.8.[16] Model (1) is estimated by OLS with bootstrapped standard errors and controlling for sub-district fixed effects. The occupational transition variables are jointly statistically significant in determining log earnings change. The occupational transitions' coefficient estimates show that individuals who were employed in non-farm activities and who remained in their initial positions all enjoyed a statistically significant gain in earnings relative to individuals who remained in farming. Conversely, those who transitioned into farming from non-farm occupations suffered statistically significant earnings losses compared to those individuals who stayed in farming. Meanwhile, all of the movements out of the farming sector result in statistically significantly positive log earnings changes. In every case, the highest point estimate for log earnings change is associated with movement into (or remaining) a non-farm employer, and is statistically significant. All of these point estimates are compared to the base case of staying in the farm sector. The conditional effects of age, education, marital status, gender, and household asset holdings all have the expected signs and are individually and jointly statistically significantly different from zero.

[16] The estimation results with absolute earnings, rather than log earnings, are qualitatively similar, as reported in Appendix Table C1.

Table 2.7: Summary statistics of variables used in the multivariate analysis

Variables	Mean	Std. Dev.	Description
Individual characteristics			All in year 2005 from SES
Age	38.48	10.86	Individual's years of age
Age^2	1598.81	871.57	Square of individual's years of age
HH head	0.41	0.49	= 1 if individual is a household head; 0 otherwise
Married	0.80	0.40	= 1 if marital status is married; 0 otherwise
Female	0.42	0.49	= 1 if gender is female; 0 otherwise
Education (Base: None/less than primary school)			
- Primary school	0.61	0.49	= 1 if completed the primary school (grade 6)
- Secondary school	0.13	0.34	= 1 if completed the secondary school (grade 9)
- High/Vocational school	0.15	0.36	= 1 if completed the high/vocational school (grade 12)
- College and above	0.07	0.26	= 1 if completed college level or higher level
Total working months	72.89	100.55	How long has individual been working in this occupation
Household characteristics			
Family size	4.26	1.72	Number of members in the household
Owned agricultural land	3.85	10.81	Agricultural area owned by household (100 Tarang-wa unit; 1 Tarang-wa = 400 m^2)
Asset Index	0.24	1.05	Asset index for household wealth based on housing characteristics, durable goods, and agricultural lands

	Farm work		NF self-employed		NF employees		NF employers	
Variables	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Individual characteristics								
Age	40.31	10.92	39.93	11.32	36.06	10.14	39.26	10.21
HH head	0.43	0.50	0.37	0.48	0.39	0.49	0.78	0.42
Married	0.84	0.37	0.84	0.37	0.75	0.43	0.90	0.30
Female	0.40	0.49	0.57	0.50	0.37	0.48	0.14	0.35
Education								
- Primary school	0.75	0.43	0.62	0.49	0.47	0.50	0.42	0.50
- Secondary school	0.10	0.30	0.16	0.36	0.15	0.36	0.15	0.37
- High/Vocational school	0.09	0.29	0.13	0.34	0.20	0.40	0.35	0.48
- College and above	0.01	0.10	0.03	0.17	0.15	0.36	0.04	0.20
Total working months	78.06	111.58	63.11	86.65	71.90	94.52	76.42	94.91
Household characteristics								
Family size	4.28	1.64	4.20	1.74	4.27	1.75	4.31	2.49
Owned agricultural land	6.62	14.84	2.22	5.07	1.86	5.92	3.35	18.17
Asset Index	-0.03	0.89	0.45	0.85	0.39	1.15	1.14	1.91

Table 2.8: Multivariate regressions of log earnings change on sector transitions and other covariates (Equation 2.1)

Dependent V: Δ ln y_{it}	(1) OLS (Bootstrap s.e.)		(2) Individual fixed effects (FE)		(3) Hausman-Taylor (random effects)	
	Coef.	Std. err.	Coef.	Std. err.	Coef.	Std. err.
Log(earning (t-1))	-0.82***	0.03	-0.91***	0.03	-0.85***	0.02
Log(earning (t-1))*t_06	0.02	0.04	0.01**	0.003	0.004**	0.001
Log(earning (t-1))*t_07	0.40***	0.03	0.01***	0.003	0.01***	0.001
Sector transitions						
Farm to (farm: base)						
NF self-employed	0.35***	0.08	0.16	0.15	0.27***	0.10
NF employee	0.66***	0.06	0.87***	0.30	0.78***	0.08
NF employer	1.05***	0.22	1.10***	0.38	1.00***	0.33
NF self-employed to						
Farm	-0.0002	0.06	-0.12	0.24	-0.08	0.08
NF self-employed	0.46***	0.05	0.37**	0.19	0.46***	0.09
NF employee	0.41***	0.06	0.62***	0.23	0.61***	0.09
NF employer	0.72***	0.09	0.29	0.20	0.47***	0.14
NF employee to						
Farm	-0.24***	0.05	0.25	0.24	0.06	0.06
NF self-employed	0.34***	0.06	0.57**	0.23	0.62***	0.09
NF employee	0.59***	0.04	0.77***	0.24	0.89***	0.08
NF employer	0.89***	0.20	1.02***	0.29	1.08***	0.25
NF employer to						
Farm	0.02	0.23	-0.06	0.27	-0.02	0.25
NF self-employed	0.54***	0.10	0.22	0.24	0.35	0.23
NF employee	0.34**	0.15	0.41*	0.24	0.48***	0.17
NF employer	0.75***	0.09	0.22	0.28	0.50*	0.28
Age	0.03***	0.01			-0.05***	0.01
Age^2	-0.0004***	0.0001			0.0005***	0.0001
HH head	0.08***	0.02			0.13***	0.03
Married	0.04	0.03			0.11***	0.04
Female	-0.05***	0.02			-0.04	0.03
Education (Less than primary school/none:base)						
Primary school	0.08	0.06			-0.31***	0.07
Secondary school	0.15**	0.06			-0.34***	0.08
High/Vocational school	0.27***	0.07			-0.20**	0.08
College and above	0.61***	0.07			0.08	0.09
Total working months	0.0003***	0.0001			0.0001	0.0001
Family size	-0.01*	0.01			-0.06***	0.01
Owned agricultural land	0.005*	0.0001			0.01*	0.004
Asset Index	0.18***	0.02			0.33***	0.02
Time effect (2006)	-0.11	0.48				
Time effect (2007)	-4.26***	0.32				

	(1) OLS (Bootstrap s.e.)		(2) Individual fixed effects (FE)		(3) Hausman-Taylor (random effects)	
Individual fixed effects	No		Yes		Random effects	
Sub-district fixed effects	Yes		No		No	
Constant	7.76***	0.37	9.18***	0.34	10.25***	0.40
Adjusted R^2	0.70		0.54			
Wald test (bootstrapped)	4027.21***				4373.30***	

***P<0.01, **P<0.05, and *P<0.10. Variance-covariance matrices are clustered at sub-district level for Model (1) and (2), and are bootstrapped for Model (1) and (3). Model (2) uses attrition weights, as described in Appendix 2A. We omit the top and the bottom one percent of the sample used, yielding 11,358 observations used.

However, other unobserved characteristics may be confounding the OLS estimates in Model (1). The five-year, four-round panel data offers the opportunity, however, to control for individual-level fixed effects so as to control for time invariant unobservables. We present those estimates as model (2). Because one might be interested in the coefficient estimates on the time invariant individual and household characteristics, model (3) presents results using a Hausman-Taylor estimator, an instrumental variables approach that enables estimation of the coefficients of time-invariant regressors while still controlling for individual-level random effects.

In the individual fixed effects model, almost all of the occupational transitions still have statistically significantly positive estimated effects on log earnings changes with an ordering in magnitude that mirrors the unconditional earnings orderings apparent in Figure 2.1. Only transitions from non-farm self-employment and employer positions into non-farm employees have greater estimated expected percentage change than those transitions into or remaining non-farm employers. However, there is no statistically significant difference between the coefficient estimates of remaining non-farm employers and transition into non-farm employee status (Table 2.9).

Qualitatively similar results emerge from model (3)'s Hausman-Taylor (H-T) estimates. The major gains come from becoming a non-farm employee and all transitions out of farming are associated with gains relative to remaining in agriculture as a primary occupation. Although the sign and statistical significance of the H-T coefficient estimates of the time-invariant observed characteristics are similar to those in OLS estimation, the sign and significance of the H-T coefficient estimates on age and education are the opposite. However, if one looks at the absolute earnings (rather than log earnings) H-T regressions (reported in Appendix Table C1), these signs on age, high school and college attainment are the same as the OLS estimators and the coefficient estimates are statistically significant. In particular, there are noticeable life cycle, gender and family size effects, while both higher individual educational attainment and greater household assets strongly and statistically significantly increase earnings.[17]

Table 2.9 presents the estimated differences in log earnings changes among occupational transitions compared within each possible past occupation, instead of remaining in farming as a base case, similar to the earnings dominance tests in Figure 2.2. The results confirm that moving to the farm sector from any non-farm occupation leads to statistically significantly lower earnings changes. Moving into the farm sector results in earnings changes 46-54 percentage points lower as compared to staying in non-farm self-employment. On average, movements into the non-farm sector increase earnings relative to remaining in farming. By contrast, shifting between non-farm

[17] As described in Appendix 2A, when we correct for non-random attrition using inverse probability weights, the coefficient estimates do not change significantly. See, in particular, Appendix Tables A5-A7.

sectors results in mixed outcomes. In the fixed effects models, for someone who is self-employed, becoming a non-farm worker increases earnings 33 percentage points relative to becoming a non-farm employer. In most cases, transitions from any of the non-farm occupations into another non-farm position lead to lower earnings growth than does staying. The lone exception is transitions from non-farm self-employment to being a non-farm employee, which is associated with a 15-25 percentage point increase in earnings, reinforcing the general impression that self-employment is less desirable than permanent salaried or wage employment. That result appears to hold even when controlling for characteristics and constraints.

Table 2.9: Change in estimated log earnings differences by sectoral transition

Model from Table 8: Sectoral transitions	(1) OLS differences	(2) FE differences	(3) H-T differences
From farm:			
to NF self-employed vs. to NF worker	-0.31***	-0.71***	-0.51***
to NF self-employed vs. to NF employer	-0.70***	-0.94**	-0.73**
to NF worker vs. to NF employer	-0.39*	-0.23	-0.21
From NF self-employed:			
to farm vs. stay NF self-employed	-0.46***	-0.50***	-0.54***
to farm vs. to NF worker	-0.41***	-0.75***	-0.69***
to farm vs. to NF employer	-0.72***	-0.42**	-0.55***
stay NF self-employed vs. to NF worker	0.05	-0.25*	-0.15*
stay NF self-employed vs. to NF employer	-0.26***	0.08	-0.01
to NF worker vs. to NF employer	-0.31***	0.33**	0.14
From NF worker:			
to farm vs. to NF self-employed	-0.58***	-0.32**	-0.56***
to farm vs. stay NF worker	-0.84***	-0.52***	-0.83***
to farm vs. to NF employer	-1.14***	-0.77***	-1.02***
to NF self-employed vs. stay NF worker	-0.25***	-0.20**	-0.27***
to NF self-employed vs. to NF employer	-0.55***	-0.45**	-0.46*
stay NF worker vs. to NF employer	-0.30	-0.25	-0.19
From NF employer:			
to farm vs. to NF self-employed	-0.52**	-0.28	-0.37
to farm vs. to NF worker	-0.32	-0.47*	-0.50*
to farm vs. stay NF employer	-0.73***	-0.28	-0.52*
to NF self-employed vs. to NF worker	0.19	-0.19	-0.13
to NF self-employed vs. stay NF employer	-0.22*	-0.01	-0.15
to NF worker vs. stay NF employer	-0.51**	0.18	-0.01

Note: A negative entry implies that the latter transition (e.g., from farm to NF worker, in the first row) yields a higher expected log earnings change than does the former (e.g., from farm to NF self-employed, in the first row); and vice versa for positive entries. *, ** and *** indicate statistically significant differences at the 10, 5 and 1 % levels, respectively.

2.4.3 Robustness checks

Although the previous regressions use individual fixed effects to control for unobserved time invariant characteristics in an attempt to disentangle the influence of occupational shifts on changes in earnings, time-varying unobservables could still drive both changes in earnings and in occupation, leading to spurious correlation that would undercut the argument that occupational transitions drive earnings gains. One important prospective class of time-varying factors unobserved in the SES data that could have such effects is village-level environmental and infrastructure variables. Improvements in village-scale infrastructure – roads, water, electricity, etc. – can change both the absolute and relative productivity of different occupations, thereby causing individual occupational transitions and hence earnings changes. Controlling for changes in infrastructure can therefore substantially obviate this prospective problem. Moreover, there might be costs associated with changing sectors and these costs (e.g, job search), are likely to decrease with the number of jobs and the rate of job growth in the local economy (Neal 1995). We therefore also control for total months spent working in the respondent's current job and changes in the ratio of total households working in particular occupations within the village. The ratios are calculated from the NRD data set to represent village employment conditions that could affect occupational switching in the village.

One approach to addressing the concern that time-varying unobservables might affect both earnings dynamics and occupational transitions is to predict the probability of these occupational movements in a first stage and then to use these predicted transition probabilities in two-stage estimation of equation (2.1). In order to do that, we have to first estimate the occupational transition probabilities using multinomial logit models, then use the predicted probabilities of occupational transition as explanatory variables in the second stage, log earnings regression. Our instruments are changes in village characteristics, reflecting changes in infrastructure and agricultural circumstances that affect the occupational choice decisions. This identification allows for more variation across villages while still allowing for variation in individual characteristics within villages. The first stage multinomial logit estimation details are discussed in Appendix 2D.

Table 2.10 reports the results of both OLS and instrumental variables regressions, with control variables from the previous survey round and annualized log earnings changes (2005-7 and 2007-10) as the dependent variable. Since we estimate each regression separately given the initial occupation in the first stage, the second stage must also be separately estimated for the three occupations besides non-farm employer. Because of the small number of observations of non-farm employers, we cannot estimate a multinomial logit for the base position of non-farm employers. The average of the predicted probabilities in each initial group is the same as the percent share reported in the transition matrix given each original occupation (Table 2.5). In each equation, each sector transition is compared to staying in the original sector. The magnitudes of the control variables' coefficient estimates and their statistical significance are similar in both the OLS and IV estimations. But although the IV estimates on occupational transitions are generally consistent in sign and magnitude with the OLS estimates and with Table 2.8's pooled estimates, and jointly statistically significant, only a few of them are individually statistically significant. This likely

Table 2.10: Estimations of log earnings change in 2005-7 and 2007-10, given initial occupation in 2005

Dependent V: $\Delta \ln y_{it}$	(1) OLS Farm		(2) IV-OLS Farm		(3) OLS NF self-employed		(4) IV-OLS NF self-employed		(5) OLS NF employee		(6) IV-OLS NF employee	
	Coef.	Std. err.	Coef.	Std. err.	Coef.	Std. err.	Coef.	Std. err.	Coef.	Std. err.	Coef.	Std. err.
Log(earnings (t-1))	-0.42***	0.04	-0.45***	0.04	-0.42***	0.02	-0.42***	0.02	-0.28***	0.02	-0.24***	0.02
Log(earning (t-1))*t_07	0.08*	0.04	0.07*	0.04	0.06***	0.02	0.06***	0.02	0.04	0.02	0.07***	0.03
Predicted probability of transitions to												
Farm	4.24***	0.47	4.30***	0.46	-0.45***	0.08	-0.45*	0.25	-0.47***	0.05	0.06	0.12
NF self-employed	0.03	0.07	-1.76***	0.65	4.73***	0.27	4.68***	0.34	-0.17***	0.05	0.14	0.29
NF employee	0.30***	0.04	0.29	0.30	-0.14**	0.06	-0.08	0.27	3.07***	0.24	2.60***	0.30
NF employer	0.65***	0.24	0.54	0.57	0.12*	0.07	-0.08	0.29	0.11	0.07	0.42	0.74
Age	-0.002	0.01	0.01	0.01	0.004	0.01	0.01	0.01	0.0002	0.01	-0.0001	0.01
Age^2	-0.0001	0.0001	-0.0001	0.0001	-0.0001	0.0001	-0.0002	0.0002	-0.00002	0.0001	-0.00004	0.0001
HH head	0.06**	0.03	0.05	0.03	0.09**	0.04	0.08*	0.04	0.03	0.02	0.02	0.02
Married	0.02	0.04	-0.05	0.05	-0.004	0.04	-0.01	0.05	0.01	0.02	-0.01	0.02
Female	0.004	0.02	0.03	0.03	-0.07**	0.03	-0.06	0.04	-0.02	0.02	-0.01	0.02
Education (Less than primary school/none:base)												
Primary school	0.10*	0.06	0.15**	0.07	-0.08	0.05	-0.06	0.07	0.01	0.04	-0.03	0.05
Secondary school	0.11	0.07	0.20**	0.08	-0.15**	0.07	-0.15*	0.08	0.07	0.05	0.03	0.06
High/Vocational school	0.15*	0.08	0.18*	0.09	-0.01	0.07	-0.02	0.09	0.19***	0.05	0.19***	0.06
College and above	0.39***	0.13	0.60***	0.16	-0.08	0.11	-0.09	0.12	0.26***	0.05	0.27***	0.06
Working months	-0.0001	0.0001	-0.0002	0.0002	-0.00001	0.0002	0.00003	0.0002	0.0003*	0.0001	0.0003*	0.0002
Family size	-0.01	0.01	-0.003	0.01	0.001	0.01	0.003	0.02	-0.01	0.01	-0.01	0.01
Owned agri land	0.01**	0.003	0.002	0.004	0.01	0.01	0.01	0.01	0.002***	0.001	0.001	0.001
Asset Index	0.08**	0.03	0.17***	0.05	0.09***	0.02	0.10***	0.03	0.05***	0.01	0.05***	0.02
Time dummy	-0.78*	0.44	-0.66	0.44	-0.71***	0.18	-0.72***	0.19	-0.42	0.27	-0.85***	0.30
Adjusted R^2	0.65		0.64		0.77		0.74		0.38		0.26	
N. of obs	2756		2756		1171		1171		2423		2423	

***P<0.01, **P<0.05, and *P<0.10. We omit the top and the bottom one percent of the sample used. Variance-covariance matrices are clustered at village level and attrition weights applied for Model (1), and are bootstrapped for Model (2).

reflects both the usual instrumental variables problem of lost efficiency and the problem of splitting the sample into smaller subsamples conditional on initial occupation, thus generating imprecise parameter estimates.[18]

As another robustness check we estimate a multinomial logit model correcting for selection bias, following Dubin and McFadden (1984) and Bourguignon et al. (2007). Occupational changes might be subject to both selection bias and endogeneity. If each group of individuals that shifts occupation differs systematically in their unobservable characteristics (e.g., skills, motivation, ability), then regression results based on individuals' observed characteristics will be biased. This method has been implemented mostly in studies of wage determinants since individuals self-select into their industry of employment. It is likely that unobservable characteristics affecting wage rates also simultaneously determine selection into the sector in which individuals work. As described in Appendix 2E, we look at occupational changes that affect earnings changes. Instead of only estimating coefficients, we calculate $E\left[\Delta \ln y \mid \hat{P}(j \text{ to } k), Z\right]$ for j,k = 0, 1, 2, 3. The estimated average earnings changes are presented in Table 11. Most of them are statistically significantly different from zero and show the expected signs, consistent with the earnings orderings manifest in the unconditional analyses described earlier. On average, shifting from farming to non-farm self-employment and non-farm employee sectors increases earnings change the

[18] When predicted probabilities are used as instruments of the actual transitions, we readily reject the null hypothesis of underidentification based on the Kleibergen Paap underidentification LM and Wald tests. Although the models are identified, only for the farmer-based equation do we reject the null with an Anderson-Rubin test for weak-identification-robust inference.

most, by 32 and 57 percentage points, respectively. However, the coefficients of the second stage regression, especially the coefficients on the selection bias correction terms are statistically insignificant (Tables E1-E3). But the results of this robustness check are consistent with the previous, individual-level fixed effects and Hausman-Taylor estimates, as well as with the unconditional earnings orderings displayed in Figure 2.2. So the core story appears robust to any of a variety of different approaches that attempt to correct for prospective statistical weaknesses in any single estimation strategy we can apply.

Table 2.11: Average log earnings changes from the selection bias correction estimations

	$E[\Delta \ln y \mid \hat{P}(j \text{ to } k), Z]$	Std Dev.
Farm to		
Farm	0.05***	0.38
NF self-employed	0.32***	0.73
NF employee	0.57***	0.73
NF self-employed to		
Farm	-0.03	1.09
NF self-employed	0.06***	0.36
NF employee	0.10***	0.66
NF employer	0.12*	0.36
NF employee to		
Farm	-0.44	0.29
NF self-employed	-0.10***	0.30
NF employee	0.03***	0.11
NF employer	0.24***	0.36

***P<0.01, **P<0.05, and *P<0.10. We omit the top and the bottom one percent of the sample used. Estimation of transition from farm to NF employer is omitted since there are only nine observations.

2.5 Conclusions

Economic growth almost always involves a transition from heavy dependence on farming to non-farm rural activity. This study reports on widespread occupational transitions in rural Thailand over a five year period, 2005-2010. Such transitions mainly involve moves into farm self-employment or non-farm employment more than into non-farm self-employment, much less into farm or non-farm employer positions. While more commercially-oriented farming with non-family employees offers demonstrably superior earnings to being a farmworker or self-employed farmer, it is still dominated by steady work as a non-farm employee or employer. The non-farm employers' and employees' earnings distributions stochastically dominate the other categories' earnings distributions, while those associated with farm workers and self-employed farmers are stochastically dominated by each of the three non-farm occupational groupings. As a result, transitions into the rural non-farm economy are associated with statistically significant earnings gains, while transitions into farming are associated with earnings losses.

But not all non-farm occupations are equally lucrative. It is more common to move down into lower-return occupations, especially non-farm self-employment and self-employed farming, than up into higher-return ones, reflecting a 'gravity effect' on the upper rungs of the occupation ladder. The vast majority of rural Thai workers appear able to quickly escape agricultural wage labor, the very lowest step on the occupational ladder, however.

Multivariate regression results, especially from our preferable model in Hausman-Taylor estimates, confirm that the biggest gains arise from becoming a non-farm employer and all transitions out of the farm sector are associated with gains relative to remaining in agriculture as a primary occupation. Moreover, both higher individual educational attainment and greater household physical capital endowments strongly and statistically significantly increase earnings, indicating the joint importance of human and physical capital, as well as climbing the occupational ladder, to earnings mobility. These results suggest that rural Thai individuals are heavily constrained in their occupational choices.

Although the most remunerative employment is as a RNFE business owner and employer, only a small number of individuals become non-farm employers, reinforcing the point that most people are not natural entrepreneurs (Kilby 1971). This result confirms similar findings from other developing and developed countries that observe far more subsistence self-employment than business owners generating paid employment for others (Mead and Liedholm 1998, Hurst and Pugsley 2011). The very small rate of transition into being a non-farm employer reflects the difficulty inherent to starting a business, while the fact that less than 40% of non-farm employers remain non-farm employers for five years – and less than 20% of farm employers remain farm employers that long – underscores the challenges of even maintaining a business with employees.

Moreover, there is a striking mismatch between at least two-thirds of rural non-farm employees working for an enterprise with ten or more employees, versus less than one percent of household enterprises employing ten or more people. Yet, most non-farm rural development programs emphasize self-employment and enterprise development, especially through micro-finance (Haggblade et al. 2007). Rural Thai households face considerable challenges in starting and maintaining, much less expanding, a farm or non-farm business and those household enterprises create few jobs. So the greatest prospects for taking advantage of the earnings gains routinely associated with occupational transitions out of farming appear to come from finding salaried or wage employment with non-household enterprises. Rural development policy might therefore aim to increase remunerative non-farm employment opportunities by established, larger-scale employers and rely less on trying to stimulate self-employment in the hopes that it will spark entrepreneurial activity and rural employment generation.

REFERENCES

Banerjee, A. & A. Newman. (1993). ”Occupational Choice and the Process of Development.” *Journal of Political Economy, 101*, 274-298.

Barrett, C. B., M. Bezuneh, D. C. Clay, & T. Reardon. (2005). “Heterogeneous constraints, incentives and income diversification strategies in rural Africa.” *Quarterly Journal of International Agriculture, 44*(1), 37-60.

Barrett, C. B., T. Reardon, & P. Webb. (2001). “Non-farm income diversification and household livelihood strategies in rural Africa: concepts, dynamics, and policy implications.” *Food Policy, 26*, 315-331.

Baulch, B. & A. Quisumbing. (2011). “Testing and adjusting for attrition in household panel data,” Chronic Poverty Research Center Toolkit Note.

Becketti, S., W. Gould, L. Lillard, & F. Welch. (1998). “The Panel Study of Income Dynamics after Fourteen Years: An Evaluation.” *Journal of Labor Economics, 6*(3), 472-92.

Bezu, S. & C. B. Barrett (2012), “Employment dynamics in the rural nonfarm sector in Ethiopia: Do the poor have time on their side?” *Journal of Development Studies*, *48*(9), 1223-1240.

Bezu, S., C.B. Barrett & S. Holden. (2012)."Does Non-farm economy offer pathways for upward mobility? Evidence from a panel data study in Ethiopia," *World Development, 40*(8), 1634-1646.

Block, S. & P. Webb (2001). “The dynamics of livelihood diversification in post-famine Ethiopia,” *Food Policy, 26*(4), 333-350.

Bourguignon, F., M. Fournier, & M. Gurgand. (2007). "Selection bias corrections based on the multinomial logit model: Monte Carlo comparisons." *Journal of Economic Surveys, 2*(1), 174-205.

Chawanote, C. (2013). "The Correlates and Dynamics of Rural Household Non-farm Business and Entrepreneurial Job Creation in Thailand" Working paper, Cornell University.

Cherdchuchai, S., & K. Otsuka. (2006). "Rural income dynamics and poverty reduction in Thai villages from 1987 to 2004." *Agricultural Economics, 35*(3), 409-423.

Cichello, P. L., G. S. Fields, & M. Leibbrandt. (2005). "Earnings and Employment Dynamics for Africans in Post-apartheid South Africa: A Panel Study of KwaZulu-Natal." *Journal of African Economies, 14*(1), 143-190.

Davidson, R., & JY. Duclos. (2000) "Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality." *Econometrica, 68*(6), 1435-1464

de Mel, S., D. McKenzie, & C. Woodruff (2008). "Who are the Microenterprise Owners?: Evidence from Sri Lanka", World Bank Working paper 4635.

Dubin, J. A. & McFadden, D. L. (1984). "An econometric analysis of residential electric appliance holdings and consumption.". *Econometrica, 52*(2), 345-362.

Fafchamps, M. (1994). "Industrial Structure and Microenterprises in Africa", *Journal of Developing Areas, 29*(1), 1-30.

Felkner, J.S. and R.M. Townsend (2011). "The Geographic Concentration of Enterprise in Developing Countries," *Quarterly Journal of Economics 126*(4), 2005-2061.

Ferreira, F. H.G., & P. Lanjouw. (2001). "Rural Non-farm Activities and Poverty in the Brazilian Northeast." *World Development, 29*(3), 509-528.

Fields, G. S. (2007). "What We Know (and Want to Know) About Earnings Mobility in Developing Countries." Working paper, Cornell University.

Fields, G. S. & M. L. Sanchez Puerta. (2005). "How Is Convergent Mobility Consistent with Rising Inequality? A Reconciliation in the Case of Argentina." Working paper, Cornell University.

Fitzgerald, J., P. Gottschalk, & R. Moffitt. (1998). "An analysis of sample attrition in panel data." *Journal of Human Resources, 33*(2), 251-299.

Foster, A. (2012). "Creating Good Employment Opportunities for the Rural Sector." *Asian Development Review* 29(1), 1-28.

Fuwa, N. (1999). "An Analysis of Social Mobility in a Village Community: The Case of a Philippine Village." *Journal of Policy Modeling, 21*(1), 101-138.

Gibson, J., and S. Olivia (2010). "The Effect of Infrastructure Access and Quality on Non-Farm Enterprises in Rural Indonesia." *World Development, 38*(5), 717-726.

Giné, X. and R.M. Townsend (2004). "Evaluation of financial liberalization: a general equilibrium model with constrained occupation choice," *Journal of Development Economics 74*(2), 269–30.

Haggblade, S., P. Hazell, and T. Reardon. (2002). “Strategies for stimulating poverty-alleviating growth in the rural non-farm economy in developing countries.” International Food and Policy Research Institute Environment and Production Technology Division Discussion Paper number 92.

Haggblade, S., D. Mead and R. Meyer (2007). “An Overview of Programs for Promoting the Rural Nonfarm Economy,” in S.Haggblade, P.B.R. Hazell, and T. Reardon, eds., *Transforming the Rural Non-farm Economy* (Baltimore: Johns Hopkins University Press).

Hazell, P. B.R., S. Haggblade, and Thomas Reardon (2007). “Structural Transformation of the Rural Non-farm Economy”, in S.Haggblade, P.B.R. Hazell, and T. Reardon, eds., *Transforming the Rural Non-farm Economy* (Baltimore: Johns Hopkins University Press).

Hung, Pham Thai, Bui Anh Tuan, and Dao Le Thanh. (2010). “Is Non-farm Diversification a Way Out of Poverty for Rural Households ? Evidence from Vietnam in 1993-2006.” PMMA working paper.

Hurst, Erik and Benjamin Wild Pugsley (2011). "What do Small Businesses Do?," *Brookings Papers on Economic Activity* 43(2), 73-142.

Isvilanonda, S., A. Ahmad, and M. Hossain. (2000). “Recent Changes in Thailand’s Rural Economy: Evidence from six villages.” *Economic and Political Weekly* December, 4644-4649.

Jonasson, E. and S. M. Helfand. (2010). “How Important are Locational Characteristics for Rural Non-agricultural Employment? Lessons from Brazil.” *World Development, 38*(5), 727-741.

Kilby, P., 1971. *Entrepreneurship and Economic Development*. New York: Free Press.

Lanjouw, P. (2001). "Non-farm Employment and Poverty in Rural El Salvador." *World Development, 29*(3), 529-547.

Liedholm, C. (2007). "Enterprise Dynamics in the Rural Non-farm Economy", in S.Haggblade, P.B.R. Hazell, and T. Reardon, eds., *Transforming the Rural Non-farm Economy* (Baltimore: Johns Hopkins University Press).

Mammen, K. and C. Paxson (2000). "Women's Work and Economic Development," *Journal of Economic Perspectives 14*(4), 141-164.

Matsumoto, T., Y. Kijima and T. Yamano. (2006). "The role of local non-farm activities and migration in reducing poverty: evidence from Ethiopia, Kenya, and Uganda", *Agricultural Economics*, *35*(3), 449-458.

Mead, D.C., Liedholm, C., 1998. The Dynamics of Micro and Small Enterprises in Developing Countries. *World Development* 26 (1), 61-74.

Mondragon-Velez, C. and X. Pena-Parga. (2008). "Business Ownership and Self-Employment in Developing Economies: The Colombian Case" in J.Lerner and A. Schoar, eds., *International Differences in Entrepreneurship* (Cambridge, MA: National Bureau for Economic Research): 89-127.

Naschold, F. and C.B. Barrett. (2011). "Do Short-Term Observed Income Changes Overstate Structural Economic Mobility?" *Oxford Bulletin of Economics and Statistics, 73*(5), 705-717.

Neal, Derek. (1995). "Industry-Specific Human Capital: Evidence from Displaced Workers", *Journal of Labor Economics, 13*(4), 653-677.

Paxson, C. (1992). “Using Weather Variability to Estimate the Response of Savings to Transitory Income in Thailand,” *American Economic Review 82*(1), 15-33.

Quadrini, V. (2000). “Entrepreneurship, Saving, and Social Mobility.” *Review of Economic Dynamics, 3*(1), 1-40.

Reardon, T. (1997), “Using Evidence of Household Income Diversification to Inform Study of the Rural Non-farm Labor Market in Africa”, *World Development, 25*(5), 735-747.

Reardon, T., J. Taylor, K. Stamoulis, P. Lanjouw and A. Balisacan. (2000). “Effect of Non-Farm Employment on Rural Income Inequality in Developing Countries: An Investment Perspective”, *Journal of Agricultural Economics, 51*(2), 266-288.

Rural Development Information Center, Community Development Department, Ministry of Interior. The University of Chicago-UTCC Research Center [distributor], 2008.

Schoar, A. (2010). “The Divide Between Subsistence and Transformational Entrepreneurship”, *Innovation Policy and the Economy, 10*(1), 57-81.

Sahn, D. E., and D. Stifel. (2003). “Exploring alternative measures of welfare in the absence of expenditure data.” *Review of Income and Wealth, 49*(4), 463-489.

Schultz, T.P. (1990). “Testing the Neoclassical Model of Family Labor Supply and Fertility,” *Journal of Human Resources 25*(4), 599-634.

Timmer, C.P. (1988). “The Agricultural Transformation.” In H. Chenery and T.N. Srinivasan eds., *Handbook of Development Economics,* Vol. 1. Amsterdam: North-Holland, pp. 275-331.

Timmer, C.P. (2002). "Agriculture and Economic Growth." In B. Gardner and G. Rausser, eds., *Handbook of Agricultural Economics, Vol. IIA*. Amsterdam: North-Holland. 1487-1546.

Woolard, I., and S. Klasen. (2005). "Determinants of Income Mobility and Household Poverty Dynamics in South Africa." *Journal of Development Studies, 41*(5), 865-897.

Wooldridge, J. M. (2002). *Economic Analysis of Cross-sectional and Panel Data*, Cambridge MA: MIT Press.

APPENDIX 2A

ATTRITION ANALYSIS

Since only 7,000 individuals, ages 15-70 years old, appear in all four SES waves while total rural individuals in the same age range each year cover more than 9,000 individuals in the initial, 2005 round, there is a potential concern that non-random attrition might bias inferences based on the balanced panel. This appendix explores issues related to attrition in this panel and explains how we have addressed those issues.

Table A1 shows the changes in rural individuals across all four waves. Only 1-2 percent of rural residents moved to urban areas from one survey round to the next. So rural outmigration is not a significant potential source of selection problems.

Table A1: Sample individuals in rural areas, ages 15-70 years

Year	Total rural	Total 2005 rural remaining	Missing	Added in	Rural to urban	Urban to rural
2005	9,153	9,153	0	0		
2006	9,415	8,443	710	972	87	58
2007	9,475	7,892	551	1,583	72	67
2010	9,984	7,000	892	2,984	196	106

We restrict attention to just those rural individuals who were employed in all four rounds, so as to avoid conflating transitions between unemployment and employment with transitions among occupations. Roughly 10-15% of the sample transitioned between unemployment and employment between rounds. As a consequence, selection bias could be an issue if there are significant differences in characteristics between those individuals present and employed in all four survey waves and those who were rural residents in 2005 but not present and employed in all four waves, given attrition – due mainly to changing household composition (due to deaths or individuals aging out of the workforce as the survey progressed) – or transitions in and out of unemployment.

Given the possibility of selection problems from restricting our analysis to those rural residents who were surveyed and employed in all four waves, we test for differences in mean characteristics between individuals in the retained subsample and those who were initially surveyed in the 2005 round. Table A2 shows that most of each group's characteristics are statistically significantly different but the key dependent variable, changes in log earnings, do not show a significant difference between the groups. As expected, those who were employed in all four waves and remained in the subsample have higher earnings on average than the full rural sample surveyed in 2005.

Since tests of mean group characteristics show significant differences, we need to test for possible attrition bias. We follow the procedure suggested by Baulch and Quisumbing (2011), first estimating an attrition probit regression, then running the tests proposed by Becketti, Gould, Lillard and Welch (1998, hereafter BGLW), then adjusting using inverse probability weights, as suggested by Fitzgerald et al. (1998) and Wooldridge (2002).

Table A2: Mean characteristics tests across groups

Variables	All rural 2005	Subsample	t-stat
Age	39.25	38.56	-3.77***
HH head	0.34	0.41	12.30***
Married	0.66	0.80	17.63***
Family size	4.31	4.26	-1.58
Education			
- Primary school	0.57	0.61	5.57***
- Secondary school	0.17	0.13	-7.39***
- High/Vocational school	0.16	0.15	-1.94*
- College and above	0.05	0.07	5.35***
Total working months	62.36	72.51	5.00***
Owned agricultural land	3338.06	3762.43	3.11***
Asset Index	0.13	0.23	5.84***
Earnings 2005	5465.97	6301.63	7.59***
Earnings 2006	5560.43	6605.51	7.94***
Earnings 2007	5582.39	6435.02	10.20***
Earnings 2010	6310.29	7191.22	8.61***
Log earnings 2005	7.76	8.05	8.94***
Log earnings 2006	7.76	8.05	10.57***
Log earnings 2007	7.94	8.16	13.37***
Log earnings 2010	8.08	8.23	8.82***
Δ Log earnings 0506	-0.04	-0.01	1.45
Δ Log earnings 0607	0.16	0.10	-2.88***
Δ Log earnings 0710	0.10	0.07	-2.01**
Δ Log earnings 0507	0.13	0.10	-1.21
Δ Log earnings 0510	0.21	0.18	-1.09

***P<0.01, **P<0.05, and *P<0.10. Mean testing uses the survey individual weight.

Table A3: Attrition probit regressions

	(1)	(2)
Pseudo-R^2	0.1527	0.2547
Wald Tests (P-value)		
Rural village attrition rate	0.000	0.000
Log (earnings) 05	0.000	
Assets	0.213	0.051
Village variables[a]	0.750	
Demography and Education[b]	0.000	0.000
Other variables[c]	0.000	0.000
No. of observations[d]	7000	9897

[a] Village variables include dummy variables for electricity available to all households in the village, insufficient water for agriculture, year-round trafficable road, agricultural loss and stagnant flood, ratio of asphalt/concrete section to total of the most convenient route to the nearest major district, travel time to the nearest district, number of soil problems, and ratios of households in the village working in establishment, manufacturing, employment, and agriculture. [b] Demography and education variables include age, age^2, household head, gender, marital status, and number of household members. [c] Other variables include demographic and education variables, asset variables, and region dummies. Variance-covariance matrices are clustered at sub-district level. [d] In both columns, we use all observations in the initial rural survey. Since there are many missing values from log(earnings) and village variables, the number of observations in column (1) has dropped as reported.

We first run a probit regression where the dependent variable takes the value of one for an individual who dropped out of the sample after the first wave (from aging out of the workforce, dying, becoming unemployed, or moving and being untraceable), then regressing that dichotomous dependent variable on baseline variables that could affect both the likelihood of attrition and the outcome variable of interest (earnings). The attrition probit regressions reported in Table A3 column 1 include individual and household characteristics, household owned agricultural land and asset index, log earnings in the initial survey round, rural village attrition rate,[19] dummy variables for each of the five main regions in Thailand, and village characteristics variables in 2005. The pseudo R^2 statistic can be interpreted as the

[19] Village attrition rate is the ratio of total individuals dropped out of the survey after the first wave to total individuals in the village in 2005.

proportion of attrition that is non-random. Wald tests are then performed to test whether observables jointly explain the predicted attrition probability. As shown in Table A3 column 1, the pseudo R^2 values are relatively low. Log earnings in 2005 and demographic variables (age, age^2, household head, female, and marital status) are statistically significantly different from zero at the 1% level, and jointly explain the attrition rate. On the other hand, village variables and asset variables insignificantly determine attrition probabilities. Many unemployed individuals have missing values of log earnings in 2005, however, as well as some missing values for village characteristics variables. If we exclude log earnings and village characteristics variables from attrition probit regressions, as shown in Table 3A column 2, we can use all the observations from the 2005 survey. This attrition probit yields a higher pseudo R^2. We can strongly reject the null hypothesis that attrition is unrelated to individual and household characteristics.

The BGLW test is based on an F-test of the joint significance of the attrition dummy and the interactions between attrition dummy and the explanatory variables when regressing the first wave outcome variable on the determinants of outcome variable, plus attrition dummy and its interactions. Hence, the model estimation is defined as

$$\Delta \ln Earnings_{0506,i} = \delta_0 + X_{05i}\delta_1 + \delta_2 Attrition_i + [X_{05i} \bullet Attrition_i]\delta_3 + v_i,$$

where X includes individual and household characteristics, household owned agricultural land and asset index, log earnings in 2005, rural village attrition rate, and regional dummy variables. The test determines whether there exist statistically significant differences between the coefficient estimates for retained and attritted individuals. The results presented in Table A4 show that the attrition dummy and its interactions are jointly statistically significant, and hence reject the null hypothesis that sample attrition is random and has no effect on coefficient estimates.[20]

Table A4: BGLW Attrition pooling tests

	(1)
R-squared	0.3967
All Interactions	
(F-stat)	4.14***
(P-value)	0.000
Attrition dummy *(P-value)*	0.260
Log (earnings) 05 *(P-value)*	0.000
A*log(earnings) 05 *(P-value)*	0.957
No. of observations[a]	6507

Variance-covariance matrices are clustered at sub-district level.
[a] Observations are dropped due to missing log earnings changes variable.

These tests suggest that non-random attrition could bias the main estimations. We use the inverse probability weighted method to correct for possible biases. Following Baulch and Quisumbig (2011), we create the ratio of predicted values from the restricted regression and unrestricted regression of reversed attrition probit where the dependent variable, RA = 1 if non-attrition, where the unrestricted regression include the same explanatory variables as the attrition probits in Table A3 column (2), while the restricted regression excludes the auxiliary variables (demographic variables, household asset variables, rural village attrition rate, and region dummies) in the first period. The inverse probability weights vary from 0.40 to 26.17 with mean 0.92.

[20] If we regress changes in log earnings between the second and the third wave on the first wave explanatory variables, we also have very low R^2 (around 0.016) and an insignificant F-test statistic on the null that all interaction terms equal zero. However, the log earnings interaction coefficient estimate remains statistically significant at the 10% level.

Tables A5 – A7 exhibit regression results that replicate the pooled OLS (with sub-district fixed effects), individual fixed effects, and Hausman-Taylor estimations, respectively, from the main body of the paper. Most of the coefficients and significance of individual coefficients are very similar. We therefore test whether coefficients with and without inverse probability weighted corrections differ statistically significantly for each of the three models. We cannot reject the null hypothesis that the parameters are equal with and without correction for attrition bias for both the pooled OLS and individual fixed effects models. A few variables' coefficient estimates are statistically significantly different for the Hausman-Taylor estimations. Therefore, although attrition bias indeed appears to exist, it seems that it does not matter much to our estimation results and resulting inferences.

Table A5: Pooled OLS regressions of log earnings change on sector transitions and other covariates

Dependent V: Δ ln y_{it}	(1) No Weight		(2) Attrition weight	
	Coef.	Std. err.	Coef.	Std. err.
Log(earning (t-1))	-0.82***	0.03	-0.79***	0.05
Log(earning (t-1))*t_06	-0.11	0.50	0.45	0.70
Log(earning (t-1))*t_07	-4.26***	0.34	-3.89***	0.50
Sector transitions				
Farm to (farm: base)				
NF self-employed	0.35***	0.08	0.37***	0.09
NF employee	0.66***	0.06	0.84***	0.15
NF employer	1.05***	0.21	1.18***	0.21
NF self-employed to				
Farm	-0.0002	0.06	0.01	0.09
NF self-employed	0.46***	0.05	0.53***	0.06
NF employee	0.41***	0.06	0.51***	0.08
NF employer	0.72***	0.10	0.85***	0.12
NF employee to				
Farm	-0.24***	0.05	-0.05	0.10
NF self-employed	0.34***	0.07	0.42***	0.09
NF employee	0.59***	0.04	0.68***	0.06
NF employer	0.89***	0.21	0.96***	0.18
NF employer to				
Farm	0.02	0.21	0.07	0.22
NF self-employed	0.54***	0.10	0.70***	0.14
NF employee	0.34**	0.15	0.36**	0.16
NF employer	0.75***	0.09	0.86***	0.10
Age	0.03***	0.01	0.03***	0.01
Age^2	-0.0004***	0.0001	-0.0004***	0.0001
HH head	0.08***	0.02	0.06***	0.02
Married	0.04	0.03	0.03	0.03
Female	-0.05**	0.02	-0.03	0.02
Education (Less than primary school/none:base)				
Primary school	0.08	0.06	0.09	0.07
Secondary school	0.15**	0.06	0.14*	0.08
High/Vocational school	0.27***	0.07	0.32***	0.09
College and above	0.61***	0.07	0.63***	0.09
Total working months	0.0003***	0.0001	0.0003**	0.0001
Family size	-0.01*	0.01	-0.01	0.01
Owned agricultural land	0.0005**	0.0002	0.0004	0.0003
Asset Index	0.18***	0.02	0.17***	0.02
Time effect (2006)	-0.11	0.50	0.45	0.70
Time effect (2007)	-4.26***	0.34	-3.89***	0.50
Constant	7.76***	0.38	7.37***	0.61
Adjusted R^2	0.70		0.71	
N. of observations	11,358		11,358	

***P<0.01, **P<0.05, and *P<0.10. Variance-covariance matrices are clustered at sub-district level. We omit the top and the bottom one percent of the sample used.

Table A6: Individual fixed effects regressions of log earnings change on sector transitions and other covariates

Dependent V: Δ ln y_{it}	(1) No Weight		(2) Attrition weight	
	Coef.	Std. err.	Coef.	Std. err.
Log(earning (t-1))	-0.93***	0.02	-0.91***	0.03
Log(earning (t-1))*t_06	0.002	0.002	0.01**	0.003
Log(earning (t-1))*t_07	0.01***	0.002	0.01***	0.003
Sector transitions				
Farm to (farm: base)				
NF self-employed	0.22***	0.08	0.16	0.15
NF employee	0.65***	0.10	0.87***	0.30
NF employer	1.00***	0.37	1.10***	0.38
NF self-employed to				
Farm	-0.09	0.09	-0.12	0.24
NF self-employed	0.32***	0.10	0.37**	0.19
NF employee	0.46***	0.11	0.62***	0.23
NF employer	0.23	0.14	0.29	0.20
NF employee to				
Farm	0.01	0.07	0.25	0.24
NF self-employed	0.47***	0.10	0.57**	0.23
NF employee	0.65***	0.10	0.77***	0.24
NF employer	0.88***	0.22	1.02***	0.29
NF employer to				
Farm	-0.14	0.23	-0.06	0.27
NF self-employed	0.13	0.20	0.22	0.24
NF employee	0.30**	0.14	0.41*	0.24
NF employer	0.13	0.25	0.22	0.28
Constant	9.63***	0.21	9.18***	0.34
Adjusted R^2	0.54		0.54	
N. of observations	11,358		11,358	

***P<0.01, **P<0.05, and *P<0.10. Variance-covariance matrices are clustered at sub-district level. We omit the top and the bottom one percent of the sample used.

Table A7: Hausman-Taylor estimation of log earnings change on sector transitions and other covariates

Dependent V: Δ ln y_{it}	(1) No Weight (bootstrapped)		(2) Attrition weight	
	Coef.	Std. err.	Coef.	Std. err.
Log(earning (t-1))	-0.85***	0.02	-0.82***	0.01
Log(earning (t-1))*t_06	0.004**	0.001	0.01***	0.002
Log(earning (t-1))*t_07	0.01***	0.001	0.01***	0.002
Sector transitions				
Farm to (farm: base)				
NF self-employed	0.27***	0.10	0.26***	0.07
NF employee	0.78***	0.08	1.06***	0.05
NF employer	1.00***	0.33	1.12***	0.23
NF self-employed to				
Farm	-0.08	0.08	-0.05	0.06
NF self-employed	0.46***	0.09	0.66***	0.07
NF employee	0.61***	0.09	0.85***	0.08
NF employer	0.47***	0.14	0.65***	0.16
NF employee to				
Farm	0.06	0.06	0.29***	0.05
NF self-employed	0.62***	0.09	0.81***	0.07
NF employee	0.89***	0.08	1.08***	0.05
NF employer	1.08***	0.25	1.13***	0.23
NF employer to				
Farm	-0.02	0.25	0.08	0.28
NF self-employed	0.35	0.23	0.56***	0.18
NF employee	0.48***	0.17	0.61**	0.25
NF employer	0.50*	0.28	0.69***	0.22
Age	-0.05***	0.01	-0.005	0.01
Age^2	0.0005***	0.0001	-0.0001	0.0001
HH head	0.13***	0.03	0.08**	0.04
Married	0.11***	0.04	0.16***	0.03
Female	-0.04	0.03	-0.04	0.03
Education (Less than primary school/none:base)				
Primary school	-0.31***	0.07	-0.27***	0.07
Secondary school	-0.34***	0.08	-0.31***	0.08
High/Vocational school	-0.20**	0.08	-0.09	0.08
College and above	0.08	0.09	0.08	0.09
Total working months	0.0001	0.0001	-0.0001	0.0001
Family size	-0.06***	0.01	-0.05***	0.01
Owned agricultural land	0.0006*	0.0003	0.001***	0.0002
Asset Index	0.33***	0.02	0.29***	0.02
Constant	10.25***	0.40	8.60***	0.17
Wald test	4373.30		24373.81	
N. of observations	11,358		11,358	

***P<0.01, **P<0.05, and *P<0.10. Variance-covariance matrices are clustered at sub-district level. We omit the top and the bottom one percent of the sample used.

APPENDIX 2B

ASSET INDEX

The asset index used as a proxy for household wealth was constructed using factor analysis based on the method of Sahn and Stifel (2003). Dummy variables for dwelling characteristics and number of durable goods owned are used in the analysis to obtain the first factor. All data come from the SES panel in 2005 (initial year). Table B1 reports summary statistics, factor loadings and scoring coefficients.

Table B1: Asset index summary statistics, factors and scoring coefficients

Variable	Mean	SD	Factor loading	Scoring coef.
Number of rooms	2.73	1.30	0.48	0.09
Housing materials (dummies)				
Cement	0.33	0.47	0.27	0.05
Wood	0.34	0.47	-0.31	-0.06
Others: local/reused materials	0.33	0.47	0.05	0.01
Having electricity	0.99	0.09	0.11	0.02
Cooking fuel (dummies)				
Gas	0.52	0.50	0.53	0.10
Others: Electricity, charcoal, wood	0.45	0.50	-0.52	-0.10
Water supply (dummies)				
Pipe	0.61	0.49	0.09	0.02
Underground water	0.35	0.48	-0.07	-0.01
Others: rain, open sources	0.04	0.19	-0.04	-0.01
Toilet: flush	0.99	0.11	0.11	0.02
Household items (number)				
Microwave	0.08	0.28	0.49	0.09
Refrigerator	0.86	0.47	0.53	0.10
Air condition	0.09	0.40	0.51	0.09
Fan	2.08	1.28	0.61	0.11
Radio	0.71	0.61	0.49	0.09
VCD-DVD player	0.66	0.57	0.56	0.10
Washing machine	0.37	0.50	0.62	0.11
Television	1.13	0.59	0.65	0.12
Cable television	0.01	0.09	0.08	0.01
Satellite dish (for TV)	0.01	0.10	0.21	0.04
Landline	0.17	0.39	0.56	0.10

Variable	Mean	SD	Factor loading	Scoring coef.
Cell phone	0.77	0.85	0.65	0.12
Computer	0.09	0.31	0.62	0.11
Internet access	0.03	0.18	0.48	0.09
Vehicles (number)				
Motorcycle	1.08	0.82	0.36	0.07
Car	0.07	0.28	0.46	0.08
Mini-truck/Van	0.19	0.46	0.45	0.08
2-wheel tractor	0.22	0.44	-0.09	-0.02
4-wheel tractor	0.01	0.11	0.10	0.02
6-wheel or higher	0.01	0.14	0.20	0.04
Owned livestock				
Buffalo	0.04	0.48	-0.07	-0.01
Cow	0.05	0.61	-0.06	-0.01

Appendix Table C1: Multivariate regressions of earnings changes (level)

Dependent V: Δy_{it} Unit: thousand baht/year	(1) OLS (Bootstrap s.e.)		(2) Individual fixed effects		(3) Hausman-Taylor (random effects)	
	Coef.	Std. err.	Coef.	Std. err.	Coef.	Std. err.
Earnings (t-1)	-0.42***	0.03	-1.20***	0.03	-0.88***	0.03
Earnings (t-1)*t_06	0.02	0.03	-0.01	0.02	-0.001	0.02
Earnings (t-1)*t_07	0.04	0.04	0.07***	0.02	0.07***	0.02
Sector transitions						
Farm to (farm: base)						
NF self-employed	12.78***	3.69	7.54**	3.65	11.57***	4.04
NF employee	23.88***	2.17	19.84***	3.49	31.01***	2.37
NF employer	58.03***	21.09	53.76**	22.27	57.47**	25.46
NF self-employed to						
Farm	-9.72***	3.24	-0.37	3.01	-2.26	3.47
NF self-employed	8.12***	1.88	17.90***	3.97	19.93***	5.14
NF employee	10.37***	3.76	19.96***	4.73	29.40***	4.18
NF employer	29.27**	13.82	34.30**	13.79	51.03***	14.31
NF employee to						
Farm	-21.28***	2.26	2.96	2.88	1.67	2.25
NF self-employed	-4.52	4.27	18.17***	5.88	17.70***	5.02
NF employee	18.05***	1.90	33.09***	3.25	49.61***	3.14
NF employer	30.49**	14.50	58.54***	13.96	60.66***	14.25
NF employer to						
Farm	-32.83*	16.84	-10.20	17.73	-3.58	19.23
NF self-employed	11.35	10.89	9.81	13.31	34.66**	16.25
NF employee	-39.11**	16.93	3.16	16.61	5.00	17.21
NF employer	26.11**	10.70	13.63	25.35	55.37**	25.77
Age	1.37***	0.26			0.80	0.68
Age^2	-0.02***	0.00			-0.01	0.01
HH head	1.93*	1.03			7.49***	2.11
Married	2.67**	1.34			8.28***	1.98
Female	-3.36***	0.87			-5.36***	1.77
Education (Less than primary school/none:base)						
Primary school	1.31	2.16			-7.17*	3.87
Secondary school	6.26**	2.65			-1.29	5.15
High/Vocational school	13.20***	2.94			10.68*	5.72
College and above	56.91***	5.61			89.05***	8.58
Total working months	0.02***	0.01			0.02***	0.00
Family size	-1.36***	0.39			-4.07***	0.60
Owned agricultural land	0.24*	0.14			0.59**	0.26
Asset Index	9.05***	1.58			23.67***	1.86
Time effect (2006)	1.51	2.11				
Time effect (2007)	5.99**	2.48				
Individual fixed effects	No		Yes		Yes	
Sub-district fixed effects	Yes		No		No	
Constant	-10.65*	6.22	62.81***	2.14	28.99	17.74
Adjusted R^2	0.21		0.57			
Wald test (bootstrapped)	1033.32***				2205.58***	

***P<0.01, **P<0.05, and *P<0.10. Variance-covariance matrices are clustering at sub-district level for Model (1) and (2), and are bootstrapped for Model (1) and (3). Model (2) uses attrition weights. We omit the top and the bottom one percent of the sample used, yielding 11,634 observations used.

APPENDIX 2D

DETERMINANTS OF OCCUPATIONAL SHIFTS

We estimate multinomial logit models to predict occupational transition probabilities and then use those predicted values in place of observed transitions in estimating equation (2.1). The multinomial logit is a form of random utility model, based on the premise that an individual compares her expected utility under different occupations and her constrained conditions:

$$U_t^e = U(O_{tj}, \ln y_{t-1}, Z, \Delta V \mid O_{t-1,j}) \tag{D1}$$

where O_{tj} is an indicator variable for occupation in period t and j = 0, 1, 2, 3 indicate farming occupation, non-farm self-employment, non-farm employee, and non-farm employer, respectively,. The variable ln y_{t-1} represents the log earnings reported in the previous year, indicating whether individuals might consider a change in occupation according to their past income draw in last period's occupation, $O_{t-1,j}$. Z denotes observed individual and household characteristics, just as in equation (2.1).

In these multinomial logit regressions we include ΔV in order to capture changes in community variables such as infrastructure and agricultural circumstances that reflect evolving environmental conditions. Table D1 enumerates these community level variables from the rural community census survey (NRD) and describes the changes in these variables for the 363 villages in the Thai SES data that match with the NRD data. Recall that the NRD data do not include 2006 and 2008. So, we study occupational transitions between 2005 and 2007 and between 2007 and 2010 based in part on changes in community variables between 2005 and 2007 and between 2007 and 2009, respectively, as a pooled panel to increase sample size.

Hence, given individuals' initial occupation, we estimate the multinomial logit:

$$\Pr(y = j \mid m) = \frac{\exp(X\Gamma(j))}{1 + \sum_{j=1}^{3} \exp(X\Gamma(j))} \tag{D2}$$

where $X\Gamma(j)_i = \gamma_0 + \gamma_1 y_{t-1,i} + Z_i\gamma_2 + \Delta V_c\gamma_3 + \nu_i$. To ensure model identification, $\Gamma(j)_i$ is set to zero when individuals stay in their previous occupation, choice *m*. In particular, if the base case is 0 = staying as a farmer, then j = 1, 2, 3 refer to the shift from farm to non-farm self-employed, the shift from farm to non-farm employee, and the shift from farm to non-farm employer, respectively. The coefficients are then interpreted with respect to staying in one's initial occupation, the base category.

The multinomial logit models are estimated separately for each initial occupation. The small number of observations of non-farm employers precludes estimation from that base position. Hence, the average marginal effects reported in Tables D2-D4 reflect only the transitions from farming, non-farm self-employment, and non-farm employee positions, respectively. The changes in village level variables jointly statistically significantly determine the transition probabilities.

As one would expect, lower initial log earnings are strongly associated with occupational shifts. People leave lower-paying jobs in search of better-paying ones. Women exhibit lower occupational mobility than men do. The main pattern we see in transitions from farming into non-farm activities is that household asset endowments are strongly and statistically significantly positively associated with transitions into non-farm employer status but negatively associated with transitions into the farming and non-farm worker categories. Those with capital are clearly more likely to take their chances starting their own enterprise. On the other hand, owned agricultural land is positively correlated with transitions into farming, confirming that land is the main factor driving work in agriculture.

Village-level variables also matter considerably. In particular, individuals living in areas lacking sufficient water for agriculture are more likely to shift from farm into non-farm employer sector. More year-round trafficable roads in the area fosters more shifts into rather than from farming. If road conditions are consistently reliable, farmers have less difficulty selling their products and hence remain on farm. Likewise, as access to electricity provided by the state increases to cover all households in a village, transitions from farm to non-farm self-employment become less likely, reflecting the importance of services to support economically viable farming.

Among those initially in the non-farm sector, fewer initial asset holdings increases the likelihood of moving into farming in the following year. Greater farm profitability in the area increases the likelihood of becoming a farmer and non-farm employer from self-employment, but increases the likelihood of becoming a non-farm worker. This reflects the linkage of agricultural prosperity to non-farm business growth.

Transportation infrastructure and employment conditions also play an important role in occupational transitions. Reduced travel time to the nearest district also induces a shift from non-farm worker to farming. More households working in establishments or manufacturing in the village also encourages transition into entrepreneurship, likely reflecting the expanded market for one’s goods or services.

Table D1: Summary statistics of changes in village characteristics

Variable	Mean	Std. Dev.	Description
Village characteristics (changes)			Changes of village-level variables between 2005 and 2007 from NRD
All HH w/ electricity	0.09	0.45	*Change in access to electricity for all households:* -1 = all households had electricity in 2005, but not all households had electricity in 2007 (worse off), 0 = same condition, 1 = better off
Insuff water for agri. (100 Rai; 1Rai = 1600 m^2)	-0.66	9.39	*Change in areas having insufficient water for agriculture:* If increasing, there were more areas in the village reporting insufficient water for agriculture.
Travel time	0.15	12.85	*Change in travel time to the nearest district (minutes):* If decreasing, it exhibits more efficient travel time from the village to the nearest district by vehicle.
Year-round well usable road	0.03	0.58	*Change in main road being usable well throughout the year:* -1 = main road was used well in 2005, but not in 2007 (worse off), 0 = same condition, 1 = better off Being usable well also implies that there is no effect from severe flooding during monsoon/rainy season, not only usable in dry season
Ratio concrete of convenient route	0.01	0.13	*Change in ratio of asphalt/concrete sections to total of the most convenient route to the nearest major district (km):* Of the total length of the most convenient route to the nearest district, more asphalt/concrete length would imply better developed infrastructure
Agriculture loss	-0.04	0.54	*Change in land utilization problem of growing plants not meeting breakeven point:* -1 = no problem in 2007, but had this problem in 2005 (better off), 0 = same condition, 1 = had problem in 2007, but no problem in 2005 (worse off)
Stagnant flood	-0.01	0.45	*Change in Stagnant flood problem:* -1 = no problem in 2007, but had this problem in 2005 (better off), 0 = same condition, 1 = had problem in 2007, but no problem in 2005 (worse off)
Soil problems	-0.08	2.13	*Change in number of soil problem*: If increasing, there were more list of soil problems in the village
Ratio HH work in establishment	-0.001	0.18	*Change in ratio of HH working in businesses/services in the village*: If increasing, there were more HH work in that sector
Ratio HH work in manufacturing	-0.01	0.09	*Change in ratio of HH working in factories/manufacturing in the village*
Ratio HH member in employment	0.01	0.27	*Change in ratio of HH whose member working as employee in the village*
Ratio HH work in agriculture	-0.001	0.36	*Change in ratio of HH working in farming*

Note: All changes in village characteristics are calculated from the NRD data set.

Table D2: Multinomial logit estimation of determinants of transitions from farm work

	Farm to NF self-employed		Farm to NF worker		Farm to NF employer	
	AME	Std. Err.	AME	Std. Err.	AME	Std. Err.
Log(earnings t-1)	-0.0119***	0.0035	-0.0124	0.0080	0.0009	0.0006
Age	0.0061*	0.0032	-0.0136*	0.0073	0.0050***	0.0017
Age^2	-0.0001	0.00004	0.0001	0.0001	-0.0001***	0.00002
HH head	-0.0074	0.0192	0.0007	0.0302	0.0040**	0.0019
Married	-0.0372	0.0344	-0.0216	0.0227	-0.0075**	0.0035
Female	0.0168	0.0148	-0.0077	0.0192	-0.0084**	0.0035
Education						
Less than primary school/none (base)						
Primary school	0.0402	0.0297	-0.0570*	0.0317	-0.0012	0.0023
Secondary school	0.0667**	0.0321	-0.0499	0.0412	-0.0264***	0.0094
High/Vocational school	0.0263	0.0366	0.0036	0.0391	0.0005	0.0033
College and above	0.1084*	0.0583	0.0311	0.0592	-0.0239**	0.0101
Total working months	-0.0001**	0.0001	-0.0001	0.0001	-0.00001	0.000004
Family size	0.0023	0.0040	0.0133**	0.0066	-0.0009	0.0007
Owned agricultural land	-0.0034***	0.0012	-0.0056*	0.0030	-0.0003**	0.0001
Asset Index	0.0434***	0.0092	-0.0459**	0.0234	0.0039***	0.0015
Change in village characteristics						
Insuff water for agri.	-0.0004	0.0004	0.0007	0.0006	0.0001*	0.0001
All HH w/ electricity	-0.0344***	0.0116	0.0169	0.0178	-0.0018	0.0014
Year-round well usable road	0.0142	0.0112	0.0054	0.0149	-0.0047**	0.0020
Ratio concrete of convenient route	0.0072	0.0465	-0.0593	0.0842	0.0005	0.0033
Travel time	-0.00005	0.0005	-0.0006	0.0007	-0.00001	0.00003
Agriculture loss	0.0142	0.0132	0.0033	0.0177	-0.0017	0.0011
Stagnant flood	-0.0037	0.0190	0.0217	0.0190	-0.0010	0.0013
Soil problems	-0.0034	0.0038	0.0082	0.0057	0.0002	0.0003
Ratio HH work in establishment	0.0415	0.0403	-0.0200	0.0533	-0.0094**	0.0039
Ratio HH work in manufacturing	-0.0150	0.0831	-0.0021	0.0934	0.0061	0.0057
Ratio HH member in employment	-0.0127	0.0222	0.1170*	0.0615	0.0027	0.0042
Ratio HH work in agriculture	0.0029	0.0186	0.0266	0.0243	-0.0024	0.0020
Time dummy	0.0236	0.0177	0.0194	0.0265	0.0064**	0.0030
Number of observations	2756					
Pseudo R^2	0.1315					

***P<0.01, **P<0.05, and *P<0.10. We omit the top and the bottom one percent of the sample used. Inverse probability weight to correct attrition bias has been applied in the estimation. Variance-covariance matrices are clustered at village level. AME = average marginal effects.

Table D3: Multinomial logit estimation of determinants of transitions from non-farm self-employment

	NF self-employed to farm		NF self-employed to NF worker		NF self-employed to NF employer	
	AME	Std. Err.	AME	Std. Err.	AME	Std. Err.
Log(earnings t-1)	-0.0268***	0.0069	-0.0116**	0.0053	0.0066*	0.0036
Age	0.0100	0.0067	-0.0007	0.0073	0.0071***	0.0022
Age^2	-0.00005	0.0001	-0.0001	0.0001	-0.0001***	0.00003
HH head	-0.0109	0.0438	0.0038	0.0217	0.0180**	0.0078
Married	-0.0043	0.0531	-0.0350	0.0257	-0.0161*	0.0096
Female	0.0070	0.0313	-0.0482**	0.0230	-0.0142*	0.0080
Education						
Less than primary school/none (base)						
Primary school	0.0710	0.0572	0.0229	0.0407	-0.0112	0.0111
Secondary school	0.0742	0.0693	-0.0098	0.0480	-0.0256*	0.0148
High/Vocational school	0.0556	0.0868	-0.0320	0.0512	-0.0023	0.0127
College and above	0.0841	0.1156	-0.0574	0.0728	0.0030	0.0214
Total working months	-0.0003**	0.0001	-0.0002	0.0001	0.00004	0.00004
Family size	0.0397***	0.0089	-0.0062	0.0059	0.0046**	0.0021
Owned agricultural land	0.0161***	0.0026	-0.0049	0.0032	-0.0003	0.0010
Asset Index	-0.0900***	0.0193	-0.0167	0.0116	-0.0019	0.0045
Change in village characteristics						
Insuff water for agri.	-0.0013	0.0018	0.0012	0.0014	-0.0001	0.0003
All HH w/ electricity	-0.0170	0.0311	-0.0212	0.0243	0.0038	0.0067
Year-round well usable road	-0.0139	0.0229	0.0155	0.0151	0.0033	0.0059
Ratio concrete of convenient route	-0.1664*	0.0990	0.0625	0.0558	0.0178	0.0190
Travel time	0.00004	0.0016	0.0010	0.0008	0.0001	0.0002
Agriculture loss	-0.0375	0.0247	0.0355**	0.0160	-0.0126	0.0079
Stagnant flood	0.0328	0.0276	0.0108	0.0205	0.0045	0.0069
Soil problems	0.0014	0.0075	-0.0091**	0.0046	0.0017	0.0014
Ratio HH work in establishment	-0.1248	0.0784	0.0118	0.0392	-0.0159	0.0162
Ratio HH work in manufacturing	0.0207	0.1362	-0.1156	0.1028	0.0824***	0.0301
Ratio HH member in employment	-0.0206	0.0661	0.0906***	0.0336	-0.0094	0.0126
Ratio HH work in agriculture	0.0012	0.0372	0.0220	0.0280	0.0038	0.0055
Time dummy	0.0342	0.0295	0.0110	0.0175	-0.0023	0.0065
Number of observations	1171					
Pseudo R^2	0.2192					

***P<0.01, **P<0.05, and *P<0.10. We omit the top and the bottom one percent of the sample used. Inverse probability weight to correct attrition bias has been applied in the estimation. Variance-covariance matrices are clustered at village level. AME = average marginal effects.

Table D4: Multinomial logit estimation of determinants of transitions from non-farm employee

	NF worker to farm		NF worker to NF self-employed		NF worker to NF employer	
	AME	Std. Err.	AME	Std. Err.	AME	Std. Err.
Log(earnings t-1)	-0.0782***	0.0141	-0.0274***	0.0097	-0.0021	0.0018
Age	0.0039	0.0052	0.0005	0.0040	0.0006	0.0011
Age^2	-0.00002	0.0001	-0.00001	0.00005	-0.00001	0.00001
HH head	-0.0118	0.0199	0.0327**	0.0140	0.0027	0.0043
Married	0.0280	0.0205	0.0153	0.0200	0.0035	0.0044
Female	-0.0156	0.0231	0.0194	0.0140	-0.0102*	0.0055
Education						
Less than primary school/none (base)						
Primary school	0.0978**	0.0464	0.0081	0.0484	0.0774***	0.0216
Secondary school	0.1418***	0.0515	-0.0099	0.0504	0.0735***	0.0209
High/Vocational school	0.0221	0.0551	-0.0182	0.0528	0.0804***	0.0222
College and above	0.0022	0.0724	-0.0774	0.0585	0.0035	0.0085
Total working months	-0.0001	0.0001	-0.0000	0.0001	-0.00004	0.00003
Family size	-0.0006	0.0064	0.0080	0.0056	0.0010	0.0009
Owned agricultural land	0.0054	0.0045	-0.0029*	0.0017	-0.0003	0.0005
Asset Index	-0.0649***	0.0133	0.0228**	0.0091	0.0032**	0.0016
Change in village characteristics						
Insuff water for agri.	0.0009	0.0016	-0.0016**	0.0007	-0.0000	0.0001
All HH w/ electricity	0.0330	0.0259	0.0091	0.0129	-0.0016	0.0024
Year-round well usable road	-0.0051	0.0184	0.0138	0.0135	0.0003	0.0027
Ratio concrete of convenient route	0.0013	0.0590	0.0133	0.0381	-0.0007	0.0095
Travel time	-0.0014***	0.0005	0.00004	0.0002	0.00004	0.0001
Agriculture loss	-0.0024	0.0172	-0.0169	0.0137	0.0012	0.0022
Stagnant flood	-0.0058	0.0212	0.0227**	0.0097	0.0011	0.0034
Soil problems	0.0005	0.0049	0.0034	0.0026	-0.0006	0.0009
Ratio HH work in establishment	-0.0774	0.0479	0.0235	0.0254	0.0118**	0.0058
Ratio HH work in manufacturing	0.5505***	0.1437	0.0196	0.0519	-0.0161	0.0101
Ratio HH member in employment	-0.0138	0.0336	-0.0131	0.0261	-0.0110***	0.0039
Ratio HH work in agriculture	0.0680***	0.0217	0.0113	0.0142	-0.0051*	0.0029
Time dummy	0.0833***	0.0172	-0.0161	0.0159	0.0021	0.0027
Number of observations	2423					
Pseudo R^2	0.1821					

***P<0.01, **P<0.05, and *P<0.10. We omit the top and the bottom one percent of the sample used. Inverse probability weight to correct attrition bias has been applied in the estimation. Variance-covariance matrices are clustered at village level. AME = average marginal effects.

APPENDIX 2E

SELECTION BIAS CORRECTIONS BASED ON THE MULTINOMIAL LOGIT MODEL

Let the changes in log earnings in the sth alternative be given by

$$\Delta \ln y_s = \alpha_s + Z_s\beta_s + u_s \tag{E1}$$

where Z contains the same covariates as in equation (2.1), including log earnings in the previous period. Let $\Delta\ln y_s$ be observed only if alternative s – an occupational shift to s given previous occupation – is chosen among four alternatives. Following the Heckman selection model, Dubin and McFadden (1984) and Bourguignon et al. (2007) include multiple correction terms to control for self-selection into the sth alternative instead of only an inverse Mills ratio term for self-selection correction. Hence, equation (E1) becomes

$$\Delta \ln y_s = \alpha_s + Z_s\beta_s + h(P_0,...,P_3) + e_s\ , \tag{E2}$$

where P_j is the probability that alternative j, j = 0, 1, 2, 3, will be chosen. P_j follows the multinomial logit model in equation (D2). Dubin and McFadden's assume that

$E(u_s) = \sigma_s \dfrac{\sqrt{6}}{\pi}\sum_{j\neq s} r_j\left[\eta_j - E(\eta_j)\right]$, where η_j is a disturbance term from equation (D1),

conditional on the alternative s being chosen. Hence, (E2) becomes

$$\Delta \ln y_s = \alpha_s + Z_s\beta_s + \sigma_s \frac{\sqrt{6}}{\pi}\left[\sum_{j\neq s} r_j\left(\frac{P_j \ln(P_j)}{1-P_j}\right) + r_s \ln(P_s)\right] + e_s \tag{E3}$$

where r_j is the correlation coefficient between disturbances u_s and η_j, and e_s is a residual with asymptotic mean zero.

Equation (E3) is estimated in two steps. The multinomial logit model for each of occupational shifts given previous occupation as in (D2) is first estimated and the predicted probabilities are substituted into the selectivity correction terms. Then, the predicted log earnings changes of each occupational shift in the second stage are averaged and reported in Table 10. Tables E1 - E3 show estimation results in the second stage of each occupation transition.

Table E1: Second stage selection bias correction estimations of transitions from farm

	Remain in Farming (F)		F to NF Self-Employment (SE)		F to Non-Farm Worker (W)	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Log(earnings$_{t-1}$)	-0.33***	0.04	-0.29**	0.12	-0.46***	0.03
Log(earning (t-1))*t_07	0.06	0.05	-0.06	0.12	0.14***	0.03
Age	0.003	0.01	-0.003	0.05	-0.004	0.01
Age^2	-0.0001	0.0001	0.0001	0.0005	0.00001	0.00
HH head	0.04	0.03	-0.13	0.13	0.01	0.04
Married	-0.02	0.05	0.02	0.22	-0.03	0.06
Female	-0.01	0.02	-0.004	0.13	-0.07*	0.04
Education						
Less than primary school/none (base)						
Primary school	-0.01	0.04	0.17	0.48	-0.13	0.09
Secondary school	0.03	0.06	0.50	0.49	-0.11	0.11
High/Vocational school	0.08	0.05	0.27	0.54	-0.04	0.10
College and above	0.23	0.17	0.54	0.81	0.08	0.15
Total working months	0.00002	0.0002	-0.0003	0.001	0.0003	0.0002
Family size	-0.02	0.01	-0.03	0.05	0.02	0.02
Owned agricultural land	0.004	0.01	-0.01	0.01	-0.005	0.005
Asset Index	0.15***	0.03	0.18	0.15	0.05	0.04
Time dummy for 2007	-0.53	0.54	0.06	1.19	-1.51***	0.25
Correction terms						
P(stay F)	-0.77	0.47	-0.01	0.96	0.39**	0.18
P(F to SE)	-0.22	0.50	-0.26	1.34	0.05	0.44
P(F to Wkr)	-0.67	0.56	0.09	1.34	0.50*	0.30
P(F to E)	-0.54	0.35	-3.73	2.32	0.50	0.56
Constant	3.43***	0.51	3.14*	1.69	5.22***	0.36
Adjusted R^2	0.39		0.47		0.81	
N. of observations	2204		209		334	

***P<0.01, **P<0.05, and *P<0.10. P(.) is the coefficient of correction terms of $P_j lnP_j/(1-P_j)$, j=1,2,3 and lnP_0. Estimation of transition from farm to NF employer is omitted since there are only nine observations.

Table E2: Second stage selection bias correction estimations of transitions from non-farm self-employment

	SE to F		Remain SE		SE to W		SE to NF Employer (E)	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Log(earnings t-1)	-0.48***	0.07	-0.34***	0.03	-0.41***	0.05	-0.16	4.17
Log(earning (t-1))*t_07	0.11**	0.05	0.05*	0.03	0.09**	0.04	-0.07	4.33
Age	0.06*	0.03	-0.01	0.01	-0.001	0.02	-0.39	1.19
Age^2	-0.001*	0.0004	-0.0001	0.0001	0.0002	0.0002	0.005	0.02
HH head	0.01	0.11	0.06*	0.03	0.01	0.08	-0.34	3.74
Married	0.18	0.14	0.01	0.04	0.08	0.07	0.51	4.41
Female	-0.16*	0.10	0.01	0.04	0.04	0.08	0.03	7.92
Education								
Less than primary school/none (base)								
Primary school	-0.16	0.24	-0.08	0.06	0.03	0.16	-0.48	3.76
Secondary school	-0.20	0.32	-0.13*	0.07	0.21	0.18	-0.41	6.37
High/Vocational school	0.01	0.25	-0.02	0.07	0.17	0.17	-0.63	3.38
College and above	-0.60	1.00	-0.06	0.11	0.33	0.27	-0.04	4.94
Total working months	-0.0001	0.0006	0.0002	0.0002	-0.0002	0.0005	0.0005	0.01
Family size	0.02	0.05	-0.01	0.01	0.002	0.03	-0.11	0.80
Owned agricultural land	0.02	0.02	-0.01	0.01	-0.001	0.02	0.01	0.73
Asset Index	0.05	0.13	0.13***	0.03	0.11	0.07	0.28	1.65
Time dummy for 2007	-1.08**	0.54	-0.60**	0.30	-0.89**	0.45	0.74	50.12
Correction terms								
P(SE to F)	0.08	0.67	0.01	0.35	0.43	0.52	-4.03	22.80
P(stay SE)	-0.16	0.22	0.16	0.18	0.20	0.14	-2.47	19.75
P(SE to Wkr)	0.24	0.52	0.04	0.28	-0.17	0.38	-3.97	41.21
P(SE to E)	-0.06	0.56	-0.15	0.32	-0.33	0.40	-4.50	31.20
Constant	3.81***	0.90	3.93***	0.37	4.02***	0.70	8.03	40.43
Adjusted R^2	0.61		0.50		0.90		0.43	
N. of observations	216		795		131		29	

***P<0.01, **P<0.05, and *P<0.10. P(.) is the coefficient of correction terms of $P_j lnP_j/(1-P_j)$, j=0,2,3 and lnP_1

Table E3: Second stage selection bias correction estimations of transitions from non-farm employee

	W to F		W to SE		Remain W		W to E	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Log(earnings t-1)	-0.44***	0.08	-0.40***	0.08	-0.20***	0.02	-0.39	0.51
Log(earning (t-1))*t_07	0.15	0.11	0.15	0.15	0.03**	0.01	0.17	0.67
Age	-0.003	0.03	-0.001	0.04	0.003	0.005	-0.05	0.18
Age^2	-0.0002	0.0004	-0.0001	0.0004	-0.00002	0.0001	0.001	0.002
HH head	-0.06	0.10	0.11	0.10	-0.001	0.01	0.08	1.54
Married	-0.02	0.10	0.10	0.11	0.02	0.01	-0.52	2.82
Female	-0.05	0.10	0.14	0.12	-0.02	0.02	-0.11	0.79
Education								
Less than primary school/none (base)								
Primary school	-0.20	0.15	-0.04	0.31	0.01	0.02	0.27	1.10
Secondary school	-0.21	0.20	0.03	0.34	0.10***	0.02		
High/Vocational school	-0.08	0.18	0.24	0.36	0.13***	0.03	0.09	0.62
College and above	-0.18	0.24	0.38	0.48	0.21***	0.03		
Total working months	-0.0003	0.0008	0.001	0.001	0.0002**	0.0001	0.002	0.004
Family size	-0.09**	0.05	0.03	0.06	-0.01	0.01	-0.03	0.82
Owned agricultural land	0.01	0.02	0.01	0.03	0.002	0.002	0.04	0.24
Asset Index	0.12	0.10	0.07	0.14	0.03**	0.02	0.09	1.35
Time dummy for 2007	-1.54	1.19	-1.59	1.73	-0.34**	0.17	-2.21	7.18
Correction terms								
P(Wkr to F)	-0.27	0.59	-0.50	0.86	0.11	0.12	-2.63	8.61
P(Wkr to SE)	0.01	0.14	-0.28	0.39	0.03	0.04	-1.50	3.12
P(stay Wkr)	0.39	0.48	-0.63	0.88	-0.06	0.10	-1.21	4.49
P(Wkr to E)	-0.57	0.57	-0.09	0.63	-0.06	0.10	-1.26	5.56
Constant	5.27***	1.23	3.76**	1.67	2.16***	0.19	5.48	7.80
Adjusted R^2	0.12		0.20		0.21		-0.25	
N. of observations	414		153		1833		21	

***P<0.01, **P<0.05, and *P<0.10. P(.) is the coefficient of correction terms of $P_j lnP_j/(1-P_j)$, j=0,1,3 and lnP_2

CHAPTER 3

THE CORRELATES AND DYNAMICS OF RURAL HOUSEHOLD NON-FARM BUSINESS AND ENTREPRENEURIAL JOB CREATION IN THAILAND

3.1 Introduction

A growing literature on entrepreneurship emphasizes the importance of the rural non-farm economy (RNFE) as a pathway out of poverty in low- and middle-income countries. Yet the growth process of microenterprises or small-and-medium enterprises (SMEs)[21] and the contribution of these enterprises to job creation have not been extensively investigated. Many empirical studies observe that most household businesses consist of one person working alone without paid, non-family employees, while most of the others are small-sized firms with limited capacity for expansion (Fafchamps 1994; Haggblade et al., 2007). While there is some business creation and expansion in any given period, there is also a contraction in firm size and the closure of others.

This study explores the entry, expansion and the contraction of rural non-farm enterprises in Thailand. It also explores the characteristics of those non-farm household entrepreneurs who expand their businesses by hiring non-family members, as well as push and pull factors that play a role in supporting job creation within the rural non-farm economy.

[21] Definitions of firm size vary somewhat across countries and studies. With the primary dataset category for firm size used in this study, the definitions of enterprise size based on number of employees are as follows: microenterprise (1-9 workers), small enterprise (10-50 workers), and medium enterprise (51-100 workers).

Subsistence self-employment could be a result of being pushed out of the agricultural sector due to natural calamities, illness and injury, or increased scarcity of land. Some people in this group might be waiting for other, better employment by a private company, nongovernmental organization (NGO), or a government agency. The non-farm sector provides them a temporary safety net to help keep them from falling into poverty and also offers an opportunity to diversify household income sources. This group might participate in the RNFE to sustain a small business with low productivity and low returns solely as a safety net, not as a growth opportunity. According to Mead and Liedholm (1998), most new rural businesses indeed begin at the least efficient and least remunerative firm size, one person, into which there are fewer barriers to entry, consistent with the view of RNFE entrepreneurial activity as largely a safety net.

In contrast, those who create or join rural non-farm businesses due to pull factors could become transformational entrepreneurs, raising their own household incomes and creating employment within their region. This group tends to be linked to high-return urban manufacturers or export-oriented subcontractors (Hazell, et. al., 2007; Liedholm, 2007). Foster and Rosenzweig (2004) point out that the rural non-farm economy is driven by the interactions between farm and rural non-farm enterprises in the labor market. Unlike new start-up firms that reflect supply-push forces, existing firm expansion creates jobs that reflect demand-pull forces and increasing efficiency. On the other hand, the reasons for RNFE business closure are mostly due to lack of product demand and shortage of working capital.

Micro and small enterprises in developing countries face several constraints, such as access to capital and skilled labor, entrepreneurial ability, and government registry requirements. Recent studies in the entrepreneurship literature find that the characteristics of those in subsistence self-employment differ from those of a potential entrepreneur who might create jobs for others, and that this is one reason why we observe a low transition rate from one-person firms to enterprises that hire non-family members (Mondragon-Velez and Pena-Parga, 2010; de Mel et al., 2008; Schoar, 2010).

Among push and pull factors for rural households to start and expand a business, one might expect capital access and thus household wealth to be the most important factor. Specifically, in the entrepreneurship literature, levels of household wealth often appear to determine the probability of becoming an entrepreneur instead of a wage worker (Evans and Jovanovic, 1989; Banerjee and Newman, 1993; Hurst and Lusardi, 2004; Paulson and Townsend, 2004). When a positive relationship between entry into entrepreneurship and wealth is found, it is often seen as evidence that borrowing constraints may impede business creation among the non-wealthy.

An alternative view is that individual entrepreneurial ability, a hard-to-measure, nontradable asset, is the key constraint to successful RNFE business creation. Buera (2009) addresses the endogenous determination of wealth that arises because an individual's ability also influences wealth level and the decision to become the entrepreneur. Those who have relatively large stocks of entrepreneurial skill tend to have a larger incentive to save up for enterprise investment (Quandrini, 2000; Buera, 2009; Toth 2011).

The objective of this study is to improve our understanding of the dynamics of enterprise and job creation in the rural non-farm economy of middle-income countries such as Thailand. The study uses the Thai Socio-Economic Survey (SES) panel data between 2005-2007 and 2010. While the short period of the panel data limits our ability to make strong causal inferences, nonetheless, predetermined household attributes and exogenous variation from rainfall and sub-district data help us to address likely endogeneity issues that necessarily complicate inference from the observed associations between household attributes and enterprise creation and growth performance.

We find wealth effects the likelihood of running or maintaining a RNFE business. We also observe far more reduction in firm size than business expansion. An ordered probit model is estimated to discover that push and pull factors around households are associated with the likelihood of operating non-farm business or hiring more workers in rural areas. The dynamics of rural non-farm business are then investigated using multinomial logit model to explore factors associated with entry, exit, and expansion of rural non-farm business.

The remainder of the paper is organized as follows. Section 2 presents an analytical framework to explain a household's allocation of labor to non-farm business and its prospective hiring of non-family members. Section 3 describes the data and variable definitions used in the empirical analysis, and offers some descriptive evidence on rural non-farm businesses in Thailand. Section 4 specifies the econometric models and estimation strategy. Section 5 discusses the results. Section 6 then provides the conclusion.

3.2 Analytical Framework

3.2.1 A model for rural household decision on operating non-farm business

We use a model that builds on familiar theories of occupational choice (Evans and Jovanovic, 1989; Banerjee and Newman, 1993; Buera, 2009; Wang, 2010; Toth, 2011; Nelson, 2011) although it will be somewhat different since the choice is not between becoming a wage employee or an entrepreneur. The rural household may diversify its activities by participating in farming or non-farm wage work, or it can start its own non-farm business. Any combination of these activities is possible, depending on how the household allocates its labor and capital. In this simple setup, we first normalize the household labor endowment to equal one, which can be allocated to work on farm (L^F), as a non-farm employee (L^w), or to start or maintain a non-farm business (L^{NF}); $L^F + L^w + L^{NF} = 1$. Household capital is used either in farm production (K^F) or in non-farm business production (K^{NF}). The household faces a credit constraint reflected in $K^F + K^{NF} \leq A + b$, where A is household total assets, and b is net borrowing, which is assumed constrained to a maximum equal to household initial wealth, K_0. If household i starts running a non-farm business in period t and hires non-family workers, l_t, then there is profit from the business, $\pi_{it+1}^{NF}(l_{it}, L_{it}^{NF}, K_{it}^{NF})$, where

$$\pi_{it+1}^{NF}(l_{it}, L_{it}^{NF}, K_{it}^{NF}) = \max_{l, L^{NF}, K^{NF}} p_{i,t+1} F(l_{it}, L_{it}^{NF}, K_{it}^{NF}) - w_{it} l_{it} - w_{it}^{H} L_{it}^{NF} - rK_{it}^{NF} \quad (3.1)$$

$$\text{s.t. } 0 \leq K_{it}^{NF} \leq A_{it} + b_{it} - K_{it}^{F}, \; b_{it} \leq K_0$$

The opportunity cost of household labor time is greater than the wage that household has to pay if it hires non-family members, $w^H > w$, in part because it must supervise hired workers.[22] Moreover, the marginal product of household members' labor, L^{NF}, is also higher than that of non-family members, l_t. One could explicitly impose this assumption on the production function of non-farm business as

$$F(l_{it}, L_{it}^{NF}, K_{it}^{NF}) = F(K_{it}^{NF}, \alpha L_{it}^{NF} + (1-\alpha) l_{it}) \tag{3.2}$$

with $\alpha > \frac{1}{2}$ such that

$$\frac{\partial F}{\partial l_{it}} = (1-\alpha) F_l(l_{it}, L_{it}^{NF}, K_{it}^{NF}) < \frac{\partial F}{\partial L_{it}^{NF}} = \alpha F_L(l_{it}, L_{it}^{NF}, K_{it}^{NF}) .$$

This assumption also reflects ideas of moral hazard for hired workers in the two-tiered labor market (Eswaran and Kotwal, 1985).

For the non-farm business, timing matters. At the beginning of period t, the business owner decides how many non-family laborers, l_t, to hire. However, the owner will receive revenue from operating the business at the end of period t +1. This happens because in the process of non-farm production, products will be produced first before getting money back from customers. The same process can be applied to other non-farm services businesses where they have to manage inventory. Hence, the business owner needs household liquid assets or access to credit in order to be able to hire non-family laborers or to hire capital. In the case of unconstrained capital, the first-order condition with respect to K_{it}^{NF} yields $\tilde{K}_{it}^{NF}$ that solves

[22] I define w^H here to differentiate wage for non-farm employees working for corporate company or government/state enterprise (permanent workers) in which they receive a higher wage rate than general workers (casual workers). In Eswaran and Kotwal (1985), there exist two-tier labor markets for permanent and casual workers where there is a moral hazard problem for hired workers. Permanent workers have more ability and do not burden employer's on-the-job supervision, hence earning higher wage. This characteristic can be viewed the same way when a household business employs family members as its labors.

$p_{t+1}F_K(K_{it}^{NF}, \alpha L_{it}^{NF} + (1-\alpha)l_{it}) = r_t$. However, if the maximum capital accessible to the non-farm business owner is constrained such that $\bar{K}_{it}^{NF} \leq A_{it} + b_{it} - K_{it}^{F}$, then $\bar{K}_{it}^{NF} < \tilde{K}_{it}^{NF}$. With diminishing marginal product of capital, we have $p_{t+1}F_K(\bar{K}_{it}^{NF}, \alpha L_{it}^{NF} + (1-\alpha)l_{it}) > r_t$ where the marginal revenue product of capital is greater than the marginal cost of capital. Consequently, we also have $p_{t+1}\alpha F_L(\bar{K}_{it}^{NF}, \alpha L_{it}^{NF} + (1-\alpha)l_{it}) < w_t^H$ and $p_{t+1}(1-\alpha)F_l(\bar{K}_{it}^{NF}, \alpha L_{it}^{NF} + (1-\alpha)l_{it}) < w_t$.

The household budget constraint will then determine whether the household operates a non-farm business or not. If the household engages in farm, non-farm employee and/or non-farm business activities, its cash, capital and labor constraints, and law of motion for assets are as follows:

$$c_{it} = \pi_{it}^{F}(L_{it}^{F}, K_{it}^{F}) + w_t^H L_t^W - w_t l_t - w_t^H L_t^{NF} + b_{it} \tag{3.3}$$

$$c_{i,t+1} = \pi_{i,t+1}^{FW}(L_{i,t+1}^{F}, K_{i,t+1}^{F}) + \pi_{i,t+1}^{NF}(l_{it}, L_{it}^{NF}, K_{it}^{NF}) - w_{i,t+1}^{H} L_{i,t+1}^{W} - w_{t+1} l_{t+1} - w_{i,t+1}^{H} L_{i,t+1}^{NF} - (1+r)b_{it} \tag{3.4}$$

$$A_{i,t+1} = (1+r)\left[A_{it} - b_{it}\right] \tag{3.5}$$

$$0 \leq K_{it}^{F} + K_{it}^{NF} \leq A_{it} + b_{it} \tag{3.6}$$

$$0 \leq L_{it}^{F} + L_{it}^{w} + L_{it}^{NF} \leq 1 \tag{3.7}$$

where c_{it} is consumption of household i at time t. Household income is composed of farm earning, wage and/or non-farm business profit, depending on household resource allocation. Thus, $L_{it}^{*F}, L_{it}^{*w}, L_{it}^{*NF} \in [0,1]$ and $K_{it}^{*F}, K_{it}^{*NF} \in [0,\ A_{it} + b_{it}]$. Note that $l_t^* > 0$ as labor supply is infinitely elastic. $l_t^* = 0$ and $L_t^{NF*} > 0$ means the household non-farm business is self-employed without any non-family employees, while $l_t^* > 0$ means household is a non-farm employer.

Suppose the household utility is a function of consumption, c_t, and separable among periods. The household's utility maximization problem is

$$\max_{c_t, L_t^F, L_t^W, L_t^{NF}, K_t^F, K_t^{NF}} u(c) = \sum_{t=0}^{\infty} \beta^t u(c_t), \tag{3.8}$$

subject to the aforementioned constraints, the household decision as to whether to operate a nonfarm business is described by

$$\max_{l_{it}, L_{it}^{NF}, K_{it}^{NF}} \left\{0,\ \pi_{it}^{NF}(l_{it}, L_{it}^{NF}, K_{it}^{NF})\right\} \tag{3.9}$$

Each period, the problem is to choose optimal $l^*(A_t, w_t, w_t^H, r_t, p_{t+1})$ and whether profits from starting/expanding business will be positive. The household thus needs to consider its endowments – labor and capital – and how to allocate them in order to maximize household utility.

Suppose there are 2 periods: $u(c_t) + \beta u(c_{t+1})$. The household decides whether to allocate its labor and capital based on the marginal returns to its resource allocation across sectors, compared to its inter-temporal marginal rate of utility by the following first order conditions:

$$\frac{u'(c_{it})}{\beta u'(c_{i,t+1})} = \frac{\partial \pi_{it}^{NF} / \partial K_{it}^{NF}}{\partial \pi_{it}^{F} / \partial K_{it}^{F}},\quad \frac{\partial \pi_{it}^{NF}}{\partial K_{it}^{NF}} = p_{t+1} F_K(K_{it}^{NF}, \alpha L_{it}^{NF} + (1-\alpha) l_{it}) - r_t$$

from $\{K_{it}^F, K_{it}^{NF}\}$ (3.10)

$$\frac{u'(c_{it})}{\beta u'(c_{i,t+1})} = \frac{\partial \pi_{it}^{F} / \partial l_{it}}{w_t},\quad \frac{\partial \pi_{it}^{F}}{\partial l_{it}} = p_{t+1}(1-\alpha) F_l(K_{it}^{NF}, \alpha L_{it}^{NF} + (1-\alpha) l_{it}) - w_t$$

from $\{l_{it}\}$ (3.11)

$$\frac{u'(c_{it})}{\beta u'(c_{i,t+1})} = \frac{\partial \pi_{it}^{NF} \big/ \partial L_{it}^{NF}}{2w_t^H}, \ \frac{\partial \pi_{it}^{NF}}{\partial L_{it}^{NF}} = p_{t+1}\alpha F_L(K_{it}^{NF}, \alpha L_{it}^{NF} + (1-\alpha)l_{it}) - w_t^H$$

from $\{L_{it}^W, L_{it}^{NF}\}$ (3.12)

$$\frac{u'(c_{it})}{\beta u'(c_{i,t+1})} = \frac{\partial \pi_{it}^{NF} \big/ \partial L_{it}^{NF}}{\partial \pi_{it}^{F} \big/ \partial L_{it}^{F} + w_t^H}$$

from $\{L_{it}^F, L_{it}^{NF}\}$ (3.13)

Suppose w_t^H increases. Given a diminishing marginal product of labor and the capital constraint mentioned earlier, in order to maintain inter-temporal equalization of the marginal utility of consumption, the household would allocate less labor to the nonfarm business. So the non-farm enterprise responds to exogenous labor market and agricultural shocks in partial equilibrium.

3.2.2 Implications for estimation

We can have a profit function represent each household's function of capital and labor, household characteristics, community characteristics, risk or shocks. From (8), we expect three scenarios for the household to decide whether to operate nonfarm business and whether to hire non-family members or just to use family members.

- If $\pi_{t+1}(l_t > 0, L_t^{NF} > 0) \geq \pi_{t+1}(l_t = 0, L_t^{NF} > 0) \geq 0$, then $l_t^* > 0$ and $L_t^{NF*} > 0$. This household operates a nonfarm business and hires non-family members.

- If $\pi_{t+1}(l_t = 0, L_t^{NF} > 0) \geq 0$, then $l_t^* = 0$ and $L_t^{NF*} > 0$. This household operates a nonfarm business, but hires no non-family members.

- If $\pi_{t+1}(L_t^{NF} > 0) < 0$, then $l_t^* = 0$ and $L_t^{NF*} = 0$. This household does not operate a nonfarm business.

This simple model assumes that the household decides simultaneously how to allocate its resources, whether to operate a non-farm business or not, and whether to hire non-family workers. Hence, we likely have endogenous regressors when using the current capital and labor allocations in estimations as these are the product of households' available choices. With the five year panel data we have, however, we can use lagged asset stocks and the ratio of household labos allocated to each sector as instruments for current capital and labor allocations in order to make sure that these variables are independent of current decision. A set of predetermined, exogenous covariates is also included as controls in the estimations, including household head's age and age squared, household head's gender and marital status, education of household head, the number of household members, village average farm earnings and non-farm wage, average provincial rate of interest, ratio of households in the village receiving credits from financial institutions and government funds, rainfall risk, and other village infrastructure and conditions.

In addition to household assets, agricultural land holdings also represent a sector-specific resource. Households possessing more agricultural land will be less likely to operate a non-farm business. However, it could be that there is a positive linkage between large farm and a non-farm business. Households with a larger farm might also run a non-farm business as a byproduct business or for additional product lines that use the same inputs.

Household head's age would be related to working experience, and the quadratic term captures life cycle effects. A married household head might be less likely to invest in a non-farm business since he/she has more family members to take care of and needs to be more careful with his/her decision to earn more stable source of incomes. However, with more family members in the household, it could be beneficial to allocate labor to a non-farm enterprise.

Agricultural earnings and non-farm wage earnings could potentially have a negative relationship with the likelihood of running a non-farm business as predicted from the model. Households enter a non-farm business when farming becomes unprofitable at the margin if the non-farm business is a source of diversification strategy of household. The wage of non-farm workers implies a higher opportunity cost of household labor. Hence, higher average wage should have negative effect on the likelihood of having a non-farm enterprise, both because household labor becomes more valuable elsewhere and because hired workers become more expensive. Finally, a higher interest rate reflects an increased cost of capital which should also have a negative correlation with the possibility of business startup.

To control for exogenous variation during the survey period, we add either variation in rainfall or rainfall shocks to the regressions. Weather shocks can be viewed as a push factor that forces rural households to withdraw labor from farming and to reallocated toward a non-farm business. Alternatively, the weather shock might itself affect a farm-related non-farm business. So weather shocks could have either a positive or negative effect on rural non-farm businesses.

3.3 Data and Descriptive Evidence

The Thai SES panel data were collected by the National Statistical Office (NSO) of Thailand in 2005 – 2007 and 2010. Beginning in 2005, NSO began tracking households and split-off individuals from sample households to create proper panel data besides the well-respected repeated cross-section SES. The panel SES survey has two main segments: i) household information on every member in the household, and ii) individual information on household members aged 15 years or older. Part one includes general information on household members, household characteristics and assets, and income from agriculture. Part two includes survey questions on education, health care, employment, incomes, expenditures, financial status (debt and savings), migration, and opinions on public policies. The survey covers every province in Thailand and randomly selects blocks of districts, sub-districts and villages, and finally selects ten households per village as in a two-stage stratified random sampling. The initial round surveyed 6,000 households with a total of 16,310 individuals (3,600 households in rural areas). Individuals were tracked in the following years' surveys. Since this study focuses on household non-farm businesses, only rural households that have members actively employed in all four survey years are selected. There are 2,101 such rural households encompassing.[23]

In the employment section of the SES, respondents were asked for their primary occupation, work status, and company size.[24] The options for primary occupation in the survey are farmer/fisherman (crops, livestock, aquaculture, fishery, hunting and gathering), production (handicrafts and basic technology manufacturing), production (industry), merchandise/own business, government/state enterprise

[23] Further explanation of the rural sample across years and attrition issues are provided in Appendix 3A.
[24] The company size represents the total number of workers including the owner. It is categorized as one worker (i.e., no employees), 2-9 workers, 10-50 workers, 51-100 workers, 101-200 workers, 201-500 workers, or over 500 workers.

employee, company/business employee, and general worker/laborer. The options for work status in the question include employer, self-employed without employees, working without pay for household business, government employee, state enterprise employee, private company employee, and cooperative group. These two questions – primary occupation and work status – are used to separate non-farm activities from farm activities at the individual level. Households where any member was operating a non-farm business either in the production or merchandise sectors are considered as having a non-farm business. If the member who operated a non-farm business employed non-family members, his/her household is classified as an employer. If his/her business recorded a total number of 1-9 non-family workers, the household is recorded as a microenterprise owner. Any business where the number of recorded workers exceeds or equals 10 is considered a small-medium enterprise (SME). On the other hand, if the members operated a non-farm business and operated alone or used only household member labor without hiring paid employees, their household business is recorded as self-employed. In some cases, if a household has multiple businesses, it is treated as one non-farm business since the survey does not have disaggregated information on each business, only individual employment and earnings information. For example, if a household has an individual who is self-employed and another individual who works as non-farm employer with five non-family member employees, then this household is classified as having a microenterprise. The rest of the households, where members only work on farms or as non-farm wage workers or both, are grouped together as households having no non-farm business.

Table 3.1 shows a transition matrix of non-farm business size as defined above, based on the number of non-family paid employees between 2005 (row) and 2010 (column). The percentage is calculated across all changes, compared to the total number of households in both periods. Almost 40 percent of rural households operated a non-farm business at some point in time during this period. The majority (about 90 percent) of rural non-farm businesses in both years were self-employed without employees and this group rarely grew larger within this period. Meanwhile, around 31 percent of households operating a non-farm business in 2005 exited the business by 2010. We also observe a tendency for microenterprise and SME firm size to contract, similar to what Mead and Liedholm (1998) observed in African countries in the 1990s.

Table 3.1: Non-farm household businesses dynamic between 2005 and 2010
(report number of households for each group and percentages of transition out of total)

			2010		
2005	No NF	SE	MIE	SME	Total
No NF	**1,283**	197	21	1	1,502
%	**61.07**	9.38	1.00	0.05	71.49
SE	183	**330**	21	5	539
%	8.71	**15.71**	1.00	0.24	25.65
MIE	13	18	**20**	1	53
%	0.62	0.86	**0.95**	0.05	2.48
SME	3	2	2	**1**	8
%	0.14	0.10	0.10	**0.05**	0.38
Total	1,482	547	64	8	2,101
%	70.54	26.04	3.05	0.38	100

No NF = No NF business, SE = NF self-employed without employee,
MIE = NF business with employees < 10 (microenterprise),
SME = NF business with employees >= 10

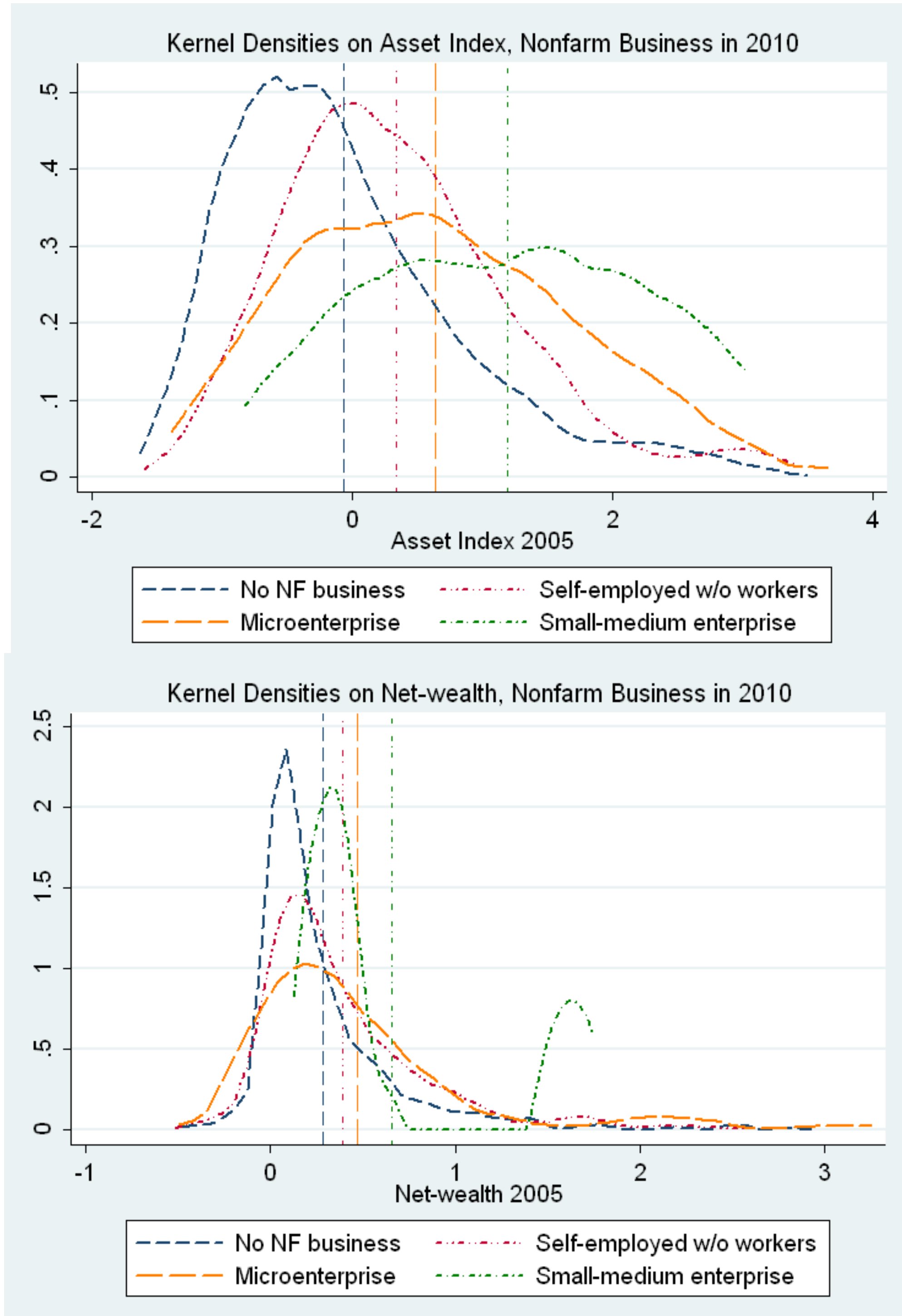
When considering the assets, wealth, savings and debt of these different groups of households, we observe some clear patterns. Since the data do not contain information on business assets, only on household and agricultural assets, an asset index using factor analysis is calculated following Sahn and Stifel (2003).[25] Household wealth each year is also calculated in 2007 Thai baht, including the value of housing and vehicles, and minus debts when we refer to net-wealth. To tackle possible endogeneity issues, authors in the entrepreneurship literature tend to use initial assets or wealth five years or more prior to the survey date (Buera, 2009; Karaivanov, 2010) to measure ex-ante wealth. However, since the data do not allow us to do so, the levels reported in 2005 are considered the initial levels of assets and wealth for each household. We cannot rule out the possibility of endogeneity, that some unobservables might co-determine both wealth and entrepreneurship ability and thus enterprise status.

Figure 3.1 shows kernel densities of the asset index and net wealth (in million baht) in 2005 by non-farm business status. The vertical line represents the mean of each distribution. It clearly shows that larger non-farm businesses have higher levels of household assets and net-wealth in the past. It also shows the same patterns if we plot the asset index distributions in 2006-2010 against non-farm business status in 2010. Although net-wealth distributions show the similar pattern to asset index

[25] The estimation details of the asset index, constructed using factor analysis following Sahn and Stifel (2003), are reported in Appendix 3B. The index includes number of rooms, housing materials, electricity, cooking fuels, water supply, toilet, number of durable goods (e.g., microwave, refrigerator, air condition, fan, television, radio, VCD-DVD player, washing machine, cable television, cell phone, landline, computer and internet), number of vehicles (motorcycles, cars, trucks, tractors), and livestock.

distributions, they do not clearly distinguish across non-farm business types and exhibit right-skewed distribution. When looking at the amount of debts, households operating a non-farm business have higher level of debts, on average.

Figures 3.2 and 3.3 show graph box distributions of saving and debt in 2005, 2007, and 2010, given non-farm enterprise status in 2010. The line in the box represents the median of the distribution while the lower and the upper ends of the box are the 25^{th} and 75^{th} percentiles, respectively. At the end of the box's arms are lower and upper adjacent values and dots above are outside values. Scatter plots of microenterprise and SMEs groups are illustrated as well since the numbers of observations in these two groups are very small. We can see that medians of saving and debts of microenterprises and SMEs are higher than those of the self-employed and the no non-farm business groups. There is not much of a change in saving over these years, but there were some changes in the distributions of debt for both microenterprises and SMEs. The debt amount on average for microenterprises and SMEs increased from 0.20 million baht in 2005 to 0.33 million baht in 2010. Compared to other households running a non-farm business, microenterprises are more likely to be burdened with debts. Although debt distributions by changes in non-farm business status between 2006 and 2010 (no business, enter, exit, and remain) exhibit similar distributions across groups and time, households who maintain a non-farm business bear a higher level of debts than the rest. On the other hand, households who had non-farm business in 2006, but did not operate in 2010, have lower level of debts (Appendix Table C1).



The vertical line represents the mean of each distribution. Attrition weights are applied.

Figure 3.1 Asset index and net-wealth distributions by non-farm enterprise status

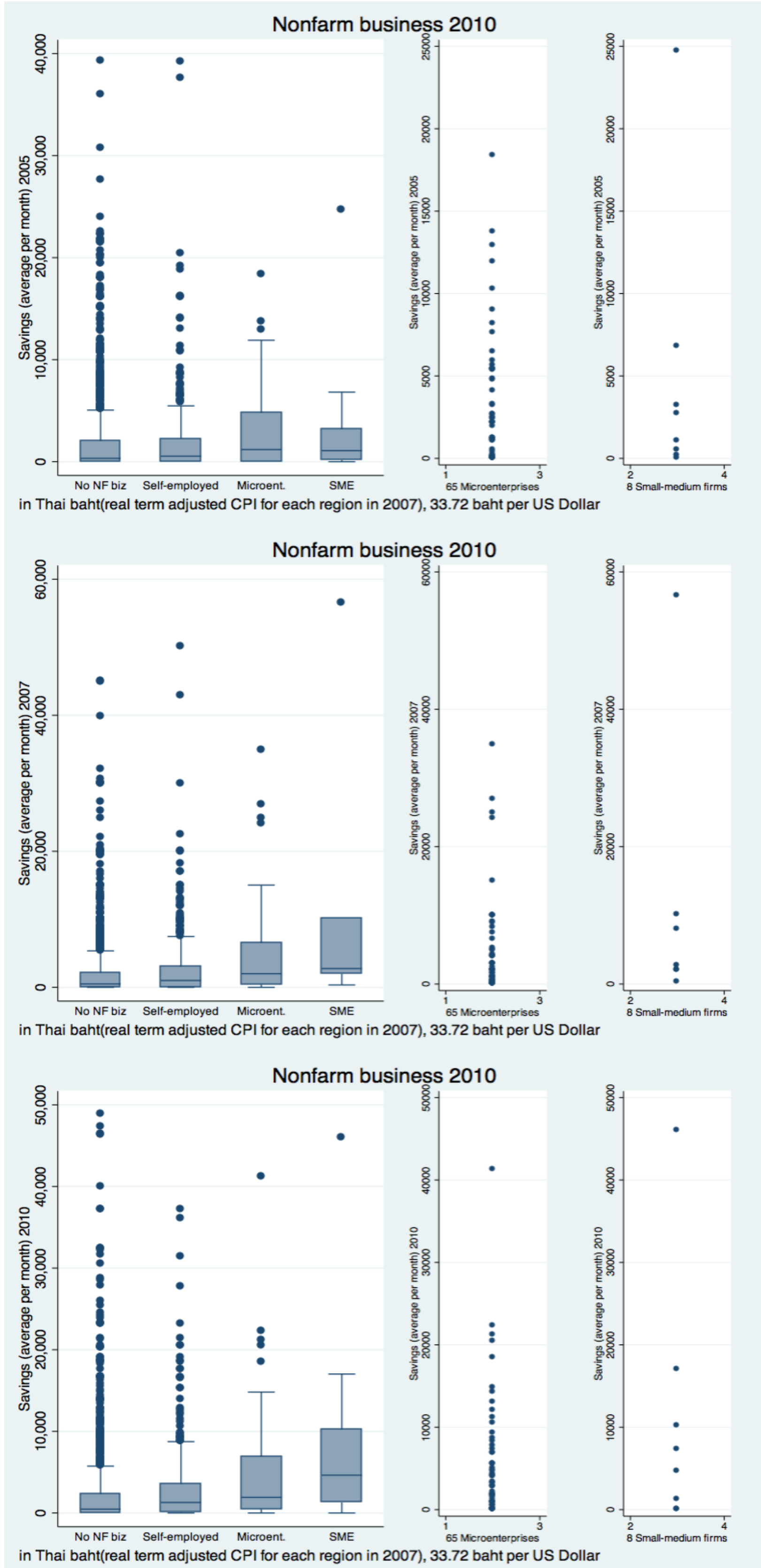


Figure 3.2: Saving distributions by nonfarm business types in 2010

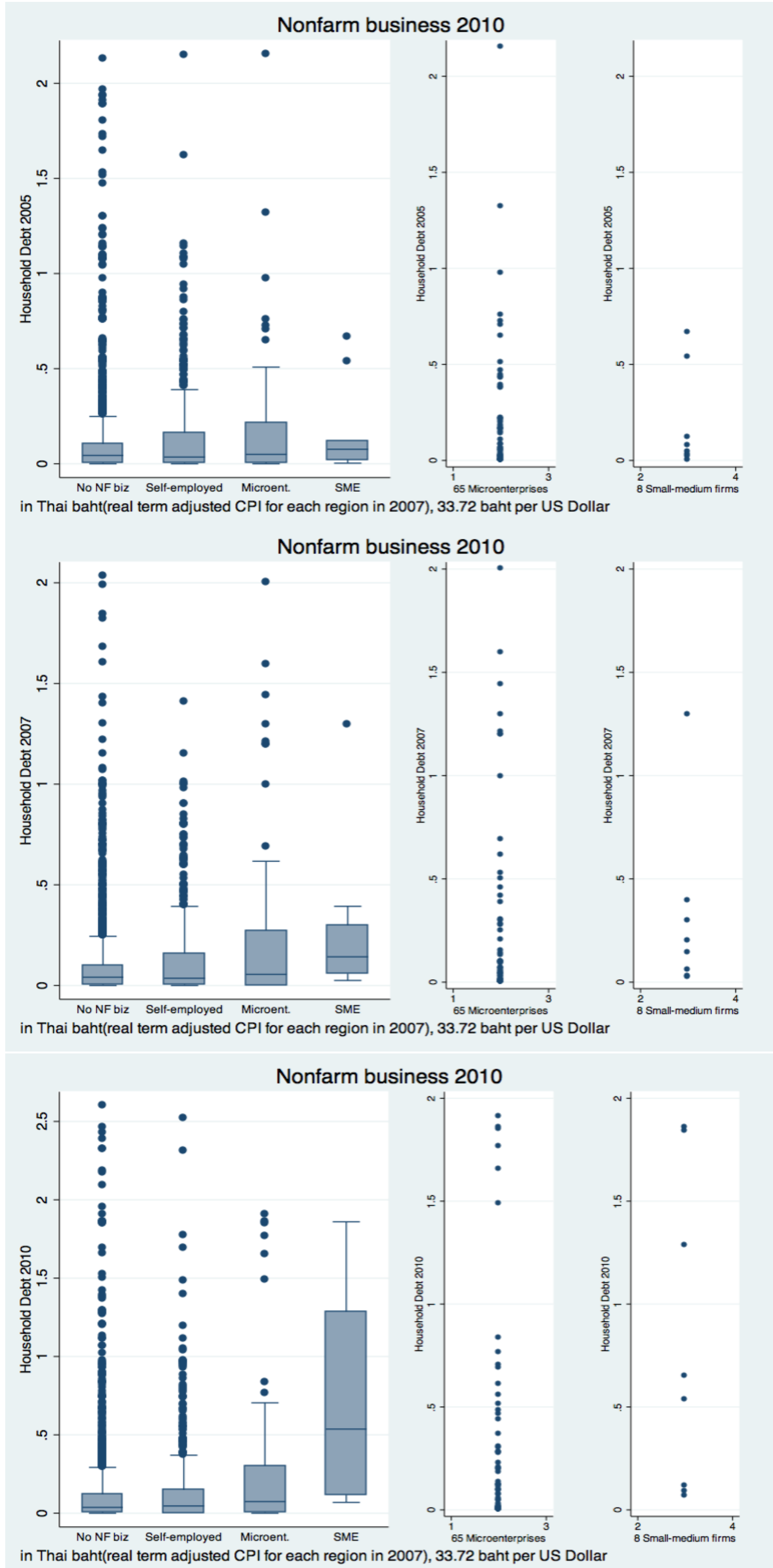


Figure 3.3: Household debt distributions by nonfarm business types in 2010

Table 3.2 shows summary statistics of the main variables used in the study, including household head and household characteristics, as well as community variables. On average, household heads in rural Thai are almost 50 years old and graduated from primary school. Most household heads are married and only 25 percent are female. More people worked on farm than worked to earn a wage or salary, but wage earnings offer higher returns. Farm earnings and wages in the study variables are calculated at community level to represent reference market rates. Interest rates are calculated as provincial level means since some communities do not have households reporting interest rates. Rural community characteristics including road conditions, travel time, soil quality, water conditions, and ratio of household receiving credits are obtained from a rural development census, the National Rural Development (NRD) data set collected at the village level by the Community Development Department of Thailand. We use only the 2005 NRD survey.

We also include monthly rainfall data from the Thai Meteorological Department for its main meteorological stations across the country (102 stations that are located closest to the sub-districts in our sample) from 1984-2009 (26 years). Thailand is located in a monsoon area where the rainy season occurs between mid-May and mid-October. We first measure variation in rainfall as the coefficient of variation of monthly rainfall between 2005 and the end of 2009 (as the 4^{th} SES wave was surveyed in January 2010). Since each part of Thailand is affected by monsoon and typhoons differently, to measure rainfall shocks, we base on the past 26 years of rainfall data. Excess rainfall and deficit rainfall are measured in terms of total yearly rainfall (in centimeter) deviations from the long run sub-district level means. Thailand averages 161.06 centimeters of rainfall per year. Excess (deficit) rainfall in year t is defined when total rainfall of year t is higher (lower) than one standard deviation from the long run sub-district mean.

Table 3.2: Summary Statistics

	Obs	Mean	Std. Dev.	Min	Max
Asset index 05	2101	0.09	1.05	-1.95	12.43
HH wealth 05	2101	0.50	0.85	0	15.56
Owned agri land (100rai)	2101	30.66	72.92	0	2400
Age HH head 05	2101	52.57	16.26	19	95
HH members 05	2101	3.71	1.56	1	12
Married 05	2101	0.76	0.43	0	1
HH head: female	2101	0.25	0.44	0	1
HH head Education 05					
educ: primary	2101	0.68	0.47	0	1
educ: secondary	2101	0.08	0.27	0	1
educ: high school	2101	0.08	0.28	0	1
educ: college/above	2101	0.04	0.19	0	1
HH farm labor ratio 2005	2101	0.29	0.33	0	1
HH NF worker ratio 2005	2101	0.23	0.28	0	1
HH NF w/o pay ratio 2005	2101	0.03	0.10	0	1
Average farm earnings 05 (10k baht)	2101	0.33	0.34	0	2.63
Average NF wage 05 (10k baht)	2101	0.78	0.71	0.01	6.22
Interest rate 05 (Provincial avg.)	2101	8.19	5.69	1	57.50
Ratio of HH receiving credits 05	2101	0.69	0.26	0	1
Excess rainfalls (ER)	2101	17.91	19.34	0	155.12
Deficit rainfalls (DR)	2101	4.75	18.41	0	127.76
CV of monthly rains during 2005-09	2101	0.17	0.12	0.04	0.95
Rural characteristics 05					
- Year-round well usable road	2067	0.54	0.50	0	1
- Ratio concrete of convenient route	2056	0.95	0.17	0	1
- Travel time to nearest city	2101	22.07	16.02	0	200
- Bad soil	2093	0.28	0.45	0	1
- Insufficient water for agri.	2093	0.45	0.50	0	1
- Stagnant flood	2093	0.17	0.38	0	1

3.4 Empirical Framework

The analytical framework of section 2 mainly explains many household decisions with respect to operating a non-farm business. In this section, we explain how we test the hypotheses implied by that model, as well as how we extend it to explore the dynamics of household non-farm enterprise operations. We estimate two different models to test the predictions of the analytical model. The first one is a non-farm business participation model, investigating the household decision to run a non-farm business or not. The second one estimates the likelihood of households having different types of non-farm business as they find the optimal number of employees. Given data limitations, we classify businesses into three size groups: self-employed (no paid employees), microenterprise (< 10 employees), and small-medium firm (≥ 10 employees). In the final, dynamic part of the analysis we explore the relationship of initial household asset/wealth to changes in household non-farm business status between 2006 and 2010, whether a household runs a non-farm business in both years, enters into non-farm business, ceases non-farm business, or stays out of non-farm business in both years.

3.4.1 Which households run rural non-farm businesses?

A binary response model enables us to identify a household's decision between running and not running a rural non-farm business. If the household decided to work as a non-farm entrepreneur, then $Y_1 = 1$; otherwise, $Y_1 = 0$. The households are assumed to maximize their earnings subject to a set of constraints as in section 2. Let V_i^* be a threshold value of utility (as a function of earnings) above which households operate a non-farm business if $v_i = V_i - V_i^* > 0$. Define $v_i = X_i'\beta + \varepsilon_i$ an index function, where X denotes the observable covariates, mainly household assets and endowments, other household characteristics, village characteristics and weather shocks. The model is represented by the following equation:

$$Y_1 = \begin{cases} 1 \text{ if } X_1'\beta + \varepsilon_1 > 0 \\ 0 \text{ if } X_1'\beta + \varepsilon_1 \leq 0 \end{cases} \tag{3.14}$$

We look at households running a non-farm business in 2010 by using household characteristics, assert index/net wealth and labor allocation in 2005 so as to obviate potential endogeneity problems. Quandrini (2000), Gentry and Hubbard (2004), and Buera (2009) mention that entrepreneur households tend to have higher ability to save, resulting in potentially higher wealth accumulation. Lagged wealth, from five years prior, is hopefully enough to disentangle wealth from current decisions. From the analytical model, farm earnings, non-farm wage, and labor allocations (represented by labor ratio in farm work, non-farm wage worker, and non-farm without pay) determine household decision on running non-farm business. All these variables are also simultaneously determined with the household decision. The use of pre-determined variables should lessen the endogeneity problem in the estimated association between all of these variables and non-farm business activity. But the coefficient estimates reported below should nonetheless not be interpreted as causal.

3.4.2 Who hires employees?

The SES survey data enable us to distinguish among non-farm businesses that are self-employed (no paid employees), microenterprise (<10 employees), and SME (≥ 10 employees). Hence, we estimate an ordered probit model specified as follows.

$$Y_2 = \begin{cases} 0 \text{ if } Y_2^* \le a_1, & \text{no non-farm business} \\ 1 \text{ if } a_1 < Y_2^* \le a_2, & \text{self-employment without employees} \\ 2 \text{ if } a_2 < Y_2^* \le a_3, & \text{microenterprise with <10 employees} \\ 3 \text{ if } a_3 < Y_2^*, & \text{small-medium sized with} \ge 10 \text{ employees} \end{cases}, \qquad (3.15)$$

where $Y_2^* = X_2'\beta_2 + \varepsilon_2$. Y_2^* represents a latent variable that determines optimal non-farm business size. As Y_2^* passes the a_1 threshold, it becomes optimal to start a self-employed non-farm business. When Y_2^* passes the a_2 (a_3) threshold, it becomes optimal to graduate to employing paid, non-family labor in a microenterprise (SME).[26]

3.4.3 Changes in non-farm business operating status

We explore the relationship between household wealth and the dynamics of household non-farm business status. To minimize endogeneity problems, we use 2005 variables to explain observed changes in non-farm business status between 2006 and 2010. Hence, given household non-farm business status in 2006, we estimate the multinomial logit:

[26] I also tried to estimate an ordered probit model using semi-nonparametric estimation. However, there are some problems with starting points (for maximum likelihood) for the same model setup used for the parametric ordered probit model. If I change the dependent variable to 2007 business status, there is no problem running the program. I will try to see if there are other ways to solve this problem.

$$\Pr(s=j)=\frac{\exp(X\Gamma(j))}{1+\sum_{j=1}^{3}\exp(X\Gamma(j))} \tag{3.16}$$

where $X\Gamma(j)_i=\gamma_0+X_i\gamma_1+\nu_i$, X_i includes household characteristics, asset index or wealth, past labor allocations, other village characteristics in 2005. j = 0, 1, 2, 3 refer to households not participating in non-farm business, entering non-farm business, exiting from non-farm business, and remaining operating in non-farm business, respectively.

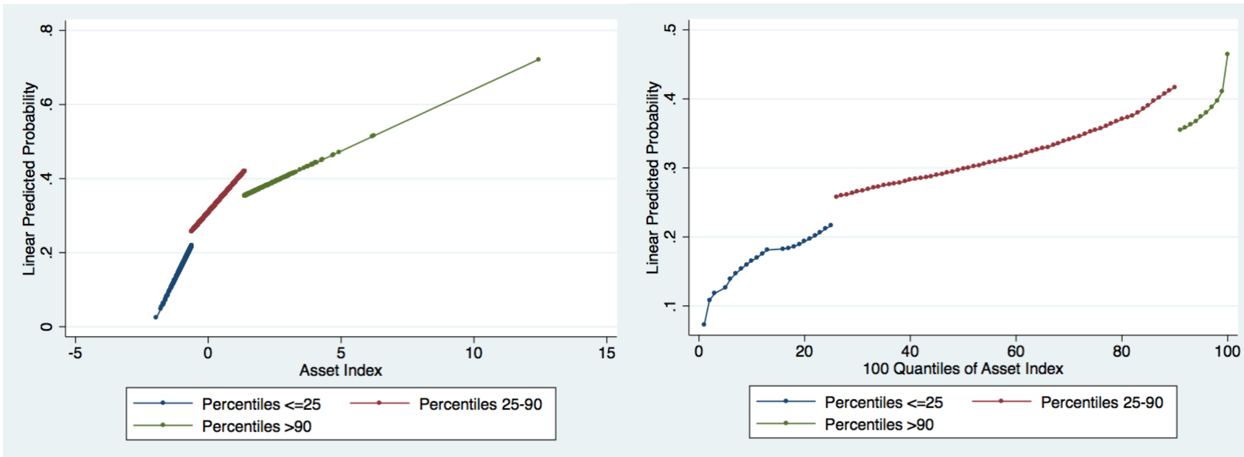
We might alternatively view non-farm business participation in 2006 and 2010 as two related successful events after conditioning on covariates. Since we do not have data during 2008-2009 to ascertain whether households ran a non-farm business or temporarily discontinued during one or more years, we can alternatively treat 2006 and 2010 as two distinct participation model observations. That suggests a bivariate probit specification allowing for correlation of the error terms from the index-function formulation of the binary outcome model. The two outcomes are determined by two unobserved latent variables (different time periods in our model) where errors of outcomes 1 (running non-farm business in 2006) and 2 (running non-farm business in 2010) are jointly normally distributed with means of 0, variances of 1, and correlations of rho, and we observe the two binary outcomes. There are four mutually exclusive outcomes: (P_1=0,P_2=0) for never having a non-farm business, (P_1=0,P_2=1) for entering into non-farm business between 2006 and 2010, (P_1=1,P_2=0) for exiting the non-farm business between 2006 and 2010, and (P_1=1,P_2=1) for having a non-farm business in both periods.

3.5 Empirical Results

3.5.1 Rural non-farm business participations

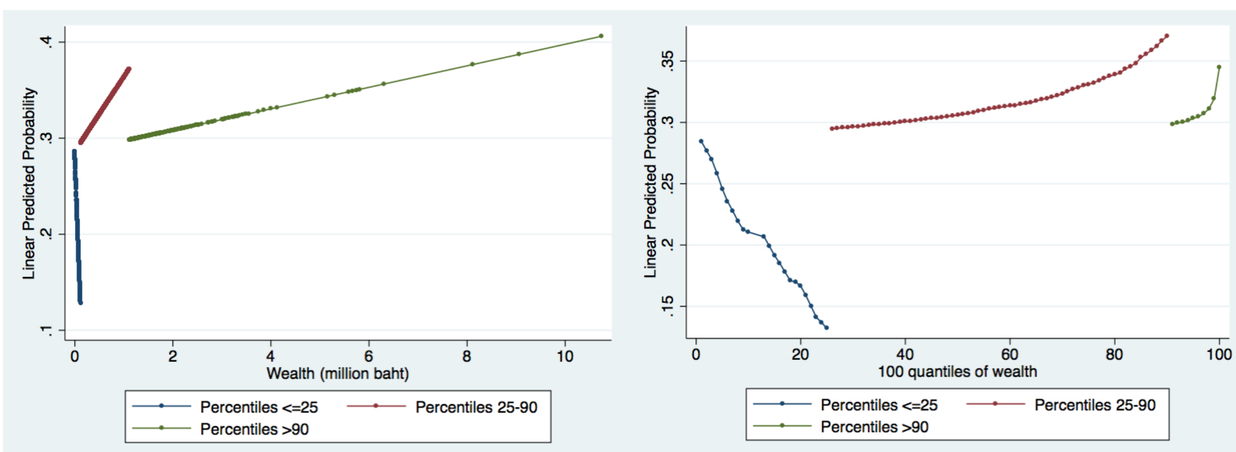
Table 3.3 reports the estimates from linear probability model (LPM) in which the dependent variable is a binary variable that takes value one if the household ran a non-farm business in 2010, and zero otherwise. I separate the asset index and wealth into three ranges of the distribution: low range for 1-25th percentiles, middle range for 25-90th percentiles, and high range for above the 90th percentiles. All three ranges are statistically significant when asset index or wealth enters the LPM in a linear relationship. Asset index positively associates with the likelihood of operating non-farm business, but the positive association falls with higher asset index values. Figure 3.4 (a) shows predicted probabilities associated with the asset index, given other variables at means while figure 3.4 (b) shows predicted probabilities on a scale of one hundred percentiles of the asset index. On the percentile scale, the predicted probability of operating a non-farm business, on average, is highest over the middle range of the asset index.

On the other hand, there is a negative relationship between wealth and the likelihood of operating a non-farm business in the lower range of wealth distribution. Figures 3.5 (a) and (b) clearly show the negatively slope relationship within the first 25th percentiles of the wealth distribution, which reflects household liquidity. It is likely that these liquid assets are used for other activities rather than running a non-farm business.



(a) Asset index (b) Percentiles of asset index

Figure 3.4: Predicted probability (LPM) of non-farm participation against asset index and asset index percentiles.



(a) Household wealth (b) Percentiles of household wealth

Figure 3.5: Predicted probability (LPM) of non-farm participation against wealth and wealth percentiles.

Table 3.3: Linear probability model estimates

Binary dependent variable: Running NF biz in 2010	(1) Coeff./se	(2) Coeff./se	(3) Coeff./se	(4) Coeff./se
Asset index 05	0.1454***	0.1091		
	(0.0243)	(0.0806)		
Asset index*(middle25-90)	-0.0636*	0.0069		
	(0.0378)	(0.0843)		
Asset index*(top90)	-0.1121***	-0.0712		
	(0.0306)	(0.0996)		
Asset index^2 05		-0.0421		
		(0.0573)		
Asset index^2*(middle25-90)		-0.0154		
		(0.0787)		
Asset index^2*(top90)		0.0389		
		(0.0584)		
Wealth 05			-1.3363***	0.0555
			(0.3641)	(1.6026)
Wealth *(middle25-90)			1.4137***	0.1995
			(0.3423)	(1.5163)
Wealth *(top90)			1.3475***	0.0282
			(0.3580)	(1.5887)
Wealth ^{2} 05				-9.8829
				(14.0401)
Wealth ^{2}*(middle25-90)				9.7444
				(13.9823)
Wealth ^{2}*(top90)				9.8726
				(14.0392)
Owned agri land	-0.0004***	-0.0004***	-0.0004**	-0.0004**
(100Tarang-wa)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
age HH head 05	0.0145***	0.0149***	0.0156***	0.0150***
	(0.0044)	(0.0045)	(0.0044)	(0.0046)
age^2	-0.0001***	-0.0001***	-0.0001***	-0.0001***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
HH members 05	-0.0321***	-0.0321***	-0.0270***	-0.0280***
	(0.0075)	(0.0075)	(0.0075)	(0.0075)
Married 05	0.0755**	0.0716**	0.0902***	0.0820**
	(0.0353)	(0.0357)	(0.0348)	(0.0351)
HH head: female	0.0619*	0.0625*	0.0678**	0.0606*
	(0.0320)	(0.0323)	(0.0321)	(0.0319)
HH head Education 05				
educ: primary	0.0162	0.0152	0.0211	0.0225
	(0.0468)	(0.0466)	(0.0473)	(0.0468)
educ: secondary	-0.0400	-0.0402	-0.0152	-0.0202
	(0.0585)	(0.0581)	(0.0588)	(0.0582)
educ: high school	-0.0304	-0.0275	-0.0018	-0.0098
	(0.0568)	(0.0568)	(0.0587)	(0.0569)
educ: college/above	-0.1641**	-0.1631**	-0.1194*	-0.1461**
	(0.0674)	(0.0681)	(0.0706)	(0.0678)
HH farm labor ratio 2005	-0.5260***	-0.5277***	-0.5619***	-0.5573***
	(0.0432)	(0.0435)	(0.0434)	(0.0436)
HH NF worker ratio 2005	-0.5200***	-0.5228***	-0.5224***	-0.5137***
	(0.0475)	(0.0480)	(0.0491)	(0.0503)
HH NF w/o pay ratio 2005	0.5512***	0.5522***	0.5732***	0.5637***
	(0.1245)	(0.1251)	(0.1262)	(0.1250)

	(1) Coeff./se	(2) Coeff./se	(3) Coeff./se	(4) Coeff./se
Average farm earnings 05	-0.0733**	-0.0719**	-0.0666**	-0.0705**
(10k baht)	(0.0295)	(0.0297)	(0.0291)	(0.0290)
Average NF wage 05	-0.0326	-0.0314	-0.0265	-0.0293
(10k baht)	(0.0213)	(0.0215)	(0.0218)	(0.0220)
Interest rate (Provincial avg.)	0.0022	0.0022	0.0021	0.0025
	(0.0017)	(0.0017)	(0.0018)	(0.0017)
Ratio of HH receiving credits	0.0946**	0.0917**	0.0925**	0.0951**
	(0.0428)	(0.0430)	(0.0436)	(0.0425)
Excess rainfalls (ER)	-0.0010*	-0.0010*	-0.0008	-0.0008
	(0.0005)	(0.0006)	(0.0005)	(0.0005)
Deficit rainfalls (DR)	0.0001	0.0001	0.0001	0.0001
	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Rural characteristics 05				
- Year-round well usable road	0.0127	0.0125	0.0111	0.0112
	(0.0234)	(0.0236)	(0.0233)	(0.0232)
- Ratio concrete of	0.0122	0.0133	0.0110	0.0117
convenient route	(0.0621)	(0.0620)	(0.0590)	(0.0585)
- Travel time to nearest city	-0.0010	-0.0009	-0.0012*	-0.0012*
	(0.0008)	(0.0008)	(0.0007)	(0.0007)
- Bad soil	-0.0605**	-0.0598**	-0.0678**	-0.0662**
	(0.0273)	(0.0274)	(0.0271)	(0.0271)
- Insuff water for agri.	0.0022	0.0020	0.0083	0.0066
	(0.0245)	(0.0245)	(0.0246)	(0.0245)
- Stagnant flood	-0.0084	-0.0109	-0.0047	-0.0034
	(0.0310)	(0.0310)	(0.0300)	(0.0299)
R2	0.246	0.245	0.236	0.238
N. of obs	2045	2045	2046	2046

***P<0.01, **P<0.05, and *P<0.10. Inverse probability weights to correct attrition bias are applied in the estimation. Variance-covariance matrices are clustered at village level.

Table 3.4 reports the estimated average marginal effects of probit estimates on the binary variable of whether the household operates a non-farm business in 2010 based on asset index or household wealth, and other covariates, in 2005. All estimates use inverse probability weights to correct for attrition as stated in Appendix 3A. In columns (1) and (2), the asset index in 2005 has a statistically significant, increasing but concave relationship with a rural household having a non-farm business in 2010. The turning point for the concave relationship is at an asset index value of 1.98 which is approximately the 95th percentile of the asset index. An increase of 100 Tarang-wa (400 square metres) statistically significantly reduces the likelihood of operating non-farm business by 0.07 percentage points, on average. The probability of having a non-farm business also increases with the age of household head, at a decreasing rate. The maximum predicted probability is at the age of 88 and 45 years old for columns (1) and (2), respectively, when holding other variables constant at their mean. If household heads graduated from college or above, their household would be less likely to operate a non-farm business. This implies that if households have a high level of education, there might be better opportunities in wage employment, depressing the probability of owning a non-farm business, ceteris paribus. More family members do not always increase the probability of running a non-farm business. If households allocate their labor to farming and/or wage work, fewer resources are left to run the non-farm business. On the other hand, when households have labor left for their own business, it increases the likelihood of operating a non-farm business. Earnings from other sources could be seen as the opportunity cost of operating non-farm business. However, if more family members work in farming or wage sector, other sources of earnings could help smooth income flow when non-farm enterprise does not do well. High earnings would tend to increase the probability of becoming an entrepreneur (Holtz-Eakin et. al, 1993). Here we have a negative sign which is more likely to imply the opportunity cost and resource competition within household, similar to what we have from the analytical framework in Section 3.2. No estimated marginal effects on wages and prices are statistically significant.

Table 3.4: Average marginal effects of probit estimation of operating nonfarm business in 2010

Binary dependent variable: Running NF biz in 2010	(1) AME/se	(2) AME/se	(3) AME/se	(4) AME/se	(5) AME/se	(6) AME/se
Asset index 05	0.1020***	0.1016***			0.0687***	0.0687***
	(0.0155)	(0.0155)			(0.0146)	(0.0145)
Asset index^2 05	-0.0258***	-0.0256***			-0.0239**	-0.0240**
	(0.0057)	(0.0057)			(0.0100)	(0.0100)
Wealth 05			0.0917***	0.0924***		
			(0.0272)	(0.0272)		
Wealth^2 05			-0.0096***	-0.0096***		
			(0.0032)	(0.0032)		
Owned agri land	-0.0007***	-0.0007***	-0.0007***	-0.0007***	-0.0004**	-0.0004*
(100Tarang-wa)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
age HH head 05	0.0177***	0.0179***	0.0198***	0.0200***	0.0050	0.0052
	(0.0053)	(0.0052)	(0.0053)	(0.0052)	(0.0046)	(0.0046)
age^2	-0.0001***	-0.0002***	-0.0002***	-0.0002***	-0.0001	-0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0000)	(0.00004)	(0.00004)
HH members 05	-0.0218***	-0.0219***	-0.0166**	-0.0167**	0.0030	0.0031
	(0.0066)	(0.0066)	(0.0067)	(0.0068)	(0.0081)	(0.0082)
Married 05	0.0698**	0.0700**	0.0752**	0.0753**	0.0229	0.0246
	(0.0345)	(0.0345)	(0.0338)	(0.0339)	(0.0311)	(0.0313)
HH head: female	0.0602**	0.0594**	0.0604**	0.0595**	0.0649**	0.0655**
	(0.0291)	(0.0289)	(0.0294)	(0.0292)	(0.0293)	(0.0292)
HH head Education 05						
educ: primary	0.0334	0.0364	0.0496	0.0524	0.0272	0.0310
	(0.0435)	(0.0436)	(0.0455)	(0.0454)	(0.0423)	(0.0420)
educ: secondary	-0.0105	-0.0052	0.0201	0.0250	0.0141	0.0180
	(0.0540)	(0.0541)	(0.0566)	(0.0564)	(0.0520)	(0.0518)
educ: high school	-0.0185	-0.0166	0.0208	0.0224	-0.0608	-0.0551
	(0.0535)	(0.0536)	(0.0552)	(0.0551)	(0.0565)	(0.0560)
educ: college/above	-0.1458**	-0.1468**	-0.1295*	-0.1304*	-0.1146	-0.1134
	(0.0676)	(0.0682)	(0.0701)	(0.0704)	(0.0731)	(0.0733)
HH farm labor ratio 2005	-0.4712***	-0.4728***	-0.5004***	-0.5016***	-0.0943*	-0.0966*
	(0.0421)	(0.0419)	(0.0428)	(0.0427)	(0.0519)	(0.0524)
HH NF worker ratio 2005	-0.4498***	-0.4503***	-0.4566***	-0.4572***	-0.0442	-0.0432
	(0.0468)	(0.0466)	(0.0486)	(0.0484)	(0.0585)	(0.0585)
HH NF w/o pay ratio 2005	0.3895***	0.3804***	0.4549***	0.4455***		
	(0.1380)	(0.1381)	(0.1366)	(0.1371)		
Average farm earnings 05	-0.0466	-0.0453	-0.0455	-0.0446	-0.0316	-0.0310
(10k baht)	(0.0314)	(0.0313)	(0.0310)	(0.0310)	(0.0275)	(0.0270)
Average NF wage 05	-0.0292	-0.0292	-0.0283	-0.0284	-0.0470**	-0.0476**
(10k baht)	(0.0196)	(0.0196)	(0.0200)	(0.0202)	(0.0204)	(0.0203)
Interest rate (Provincial avg.)	0.0012	0.0017	0.0019	0.0024	0.0012	0.0015
	(0.0015)	(0.0015)	(0.0015)	(0.0015)	(0.0016)	(0.0016)
Ratio of HH receiving credits	0.0837**	0.0919**	0.0909**	0.0979**	0.1170***	0.1226***
	(0.0405)	(0.0404)	(0.0410)	(0.0408)	(0.0381)	(0.0379)
Excess rainfalls (ER)	-0.0012**		-0.0011*		-0.0008	
	(0.0005)		(0.0005)		(0.0005)	
Deficit rainfalls (DR)	-0.0000		0.0001		0.0000	
	(0.0004)		(0.0004)		(0.0005)	
CV of monthly rains during		-0.1009		-0.0850		-0.1010
2005-09		(0.0811)		(0.0810)		(0.0901)

	(1) AME/se	(2) AME/se	(3) AME/se	(4) AME/se	(5) AME/se	(6) AME/se
Rural characteristics 05						
- Year-round well usable road	0.0080	0.0106	0.0141	0.0165	0.0264	0.0284
	(0.0223)	(0.0223)	(0.0221)	(0.0221)	(0.0198)	(0.0198)
- Ratio concrete of	0.0059	0.0111	0.0133	0.0176	-0.0156	-0.0136
convenient route	(0.0607)	(0.0584)	(0.0569)	(0.0548)	(0.0516)	(0.0499)
- Travel time to nearest city	-0.0008	-0.0008	-0.0010	-0.0010	-0.0006	-0.0006
	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0007)
- Bad soil	-0.0572**	-0.0532*	-0.0617**	-0.0582**	-0.0417*	-0.0396*
	(0.0269)	(0.0272)	(0.0268)	(0.0271)	(0.0238)	(0.0239)
- Insuff water for agri.	-0.0027	-0.0037	0.0061	0.0054	0.0401*	0.0399*
	(0.0233)	(0.0232)	(0.0232)	(0.0232)	(0.0209)	(0.0208)
- Stagnant flood	-0.0117	-0.0084	-0.0051	-0.0023	-0.0266	-0.0244
	(0.0311)	(0.0312)	(0.0299)	(0.0302)	(0.0288)	(0.0289)
Pseudo R2	0.2317	0.2311	0.2162	0.2148	0.0910	0.0898
N. of obs	2048	2048	2048	2048	1455	1455

***P<0.01, **P<0.05, and *P<0.10. Inverse probability weights to correct attrition bias are applied in the estimation. Variance-covariance matrices are clustered at village level. Other variables include provincial dummies. Columns (1) – (4) contain all sample while columns (5) and (6) include only households that do not have non-farm business in 2005.

Weather risk from rainfall shocks does not have a strong significant effect overall. Column (1) uses total excess and deficit rainfalls during 2005-2009, and shows that extreme weather has a negative correlation with the probability of having a non-farm business. When we use coefficient of variation in column (2), it also shows that areas with high weather risk have a lower likelihood of having a non-farm business although the estimated effect is statistically insignificant. This negative relationship could imply non-farm business linkages with the agricultural sector. We also interact rainfall variables with asset/wealth, agricultural land holdings, and earnings variables. However, these interactions are not statistically significant.

Rural infrastructure and related community-level conditions show the expected associations, but the estimates are not statistically significant. A usable road with good year-round condition and quality road increase the likelihood of operating a non-farm business. More travel time to the nearest city reduces that likelihood. Bad soil quality and the presence of water problems lower the possibility of operating a non-farm business, but only statistically significantly in the case of soil quality.

Columns (3) and (4) use the household wealth level instead of the estimated asset index.[27] The likelihood of having a non-farm business statistically significantly increases with wealth, increasing at a decreasing rate. The estimated turning point is at wealth equal to 4.8 million baht, which is at the 100th percentile of wealth distribution. The rest of the results when using household wealth are likewise similar to those of columns (1) and (2).

Columns (5) and (6) use only observations that reported no non-farm business in 2005, hence the smaller number of observations.[28] Overall, the results are similar to columns (1) and (2), with less statistically significant point estimates, due quite possibly to the smaller sample size. Conditional on no non-farm business in 2005, the non-farm wage is now strongly negatively associated with the likelihood of starting a non-farm business by 2010. The non-farm wage could be crucial to household decisions as to whether to allocate labor to starting a non-farm business or not. Moreover, the estimated average marginal effects of credit availability in the community on starting non-farm business are higher for households without a non-farm business in 2005 than for the full sample, which seems to signal liquidity

[27] If we use household net-wealth, the average marginal effects are not statistically significant. We see from data descriptive analysis in Section 3 that net wealth distributions are not much different across non-farm business groups, but that household wealth and debts are different.
[28] Columns (1)-(4) pay attention only to the non-farm business status in 2010; non-farm business status in 2005 does not matter. Columns (5) and (6) select only households with “no non-farm business” status in 2005 so as to see if these households have started a non-farm business by 2010.

constraints. Insufficient water agriculture also increased the likelihood of household starting non-farm business in 2010.

3.5.2 Ordered probit estimations of non-farm business types

Table 3.5 shows the estimated average marginal effects of the ordered probit estimations of household decision in running non-farm business based on firm size. As discussed previously, we have four ordered firm size groups, with no non-farm business as a base, non-farm business without employees, then microenterprise (1-9 employees), and then SME (10 or more employees). Each column within the panel presents the estimated average marginal effects for each type of non-farm business in 2010, including no non-farm business. Panel (1) uses rainfall shocks while panel (2) uses the coefficient of variation of rainfall. The estimated average marginal effects on the likelihood of having non-farm business show the same signs with smaller magnitude for higher household business size since we have a smaller number of observations associated with those larger businesses. The signs of the base case's average marginal effects are all opposite to the signs of the higher non-farm business size. Most of the statistically significant estimated average marginal effects show similar results as when we estimate probit models of non-farm business participation.

Assets are statistically significantly related to who runs a non-farm business and hires more non-family labors. The turning point for the concave relationship (convex case for no non-farm business), holding other covariates constant, is at asset index equal 2.53 for the first three orders, and 2.44 for SME, all of which all at approximately the 97th percentile of the asset index. So the likelihood of operating a non-farm business of any size is effectively increasing throughout the asset holding distribution.

In the ordered probit model, however, two additional variables show statistical significance, compared to the probit model. First, households in a higher average wage earnings area are less likely to operate a non-farm business in 2010. A ten thousand baht increase in the average non-farm wage reduces the probability of having a non-farm self-employed business by 2.9 percentage points, on average. The other additional significant variable is excess rainfall. More rainfall increases the likelihood of having no non-farm business. Likewise, households in the areas with more exposure to rainfall risk are less likely to operate a non-farm business. Based on these results, non-farm business does not seem to be a way to secure household earnings from weather risk exposure.

Coefficient estimates are reported in Appendix Table C2. The latent non-farm enterprise size variable is increasing with the asset index at a decreasing rate and decreasing with agricultural land holdings. The threshold parameters are statistically significantly different from each other. A likelihood ratio test of interaction terms between rainfall shocks (or cv) and asset index, agricultural land holdings and average earnings does not support the unrestricted model as opposed to restricted model without interactions with rainfall shocks. Figure 3.6 illustrates the predicted probabilities of each non-farm business group plotted against the household asset index. The predicted probabilities of operating a non-farm SME increase with the asset index while the predicted probabilities of having no non-farm business show the opposite pattern. The predicted probabilities of the other two cases show an inverted-U shape.

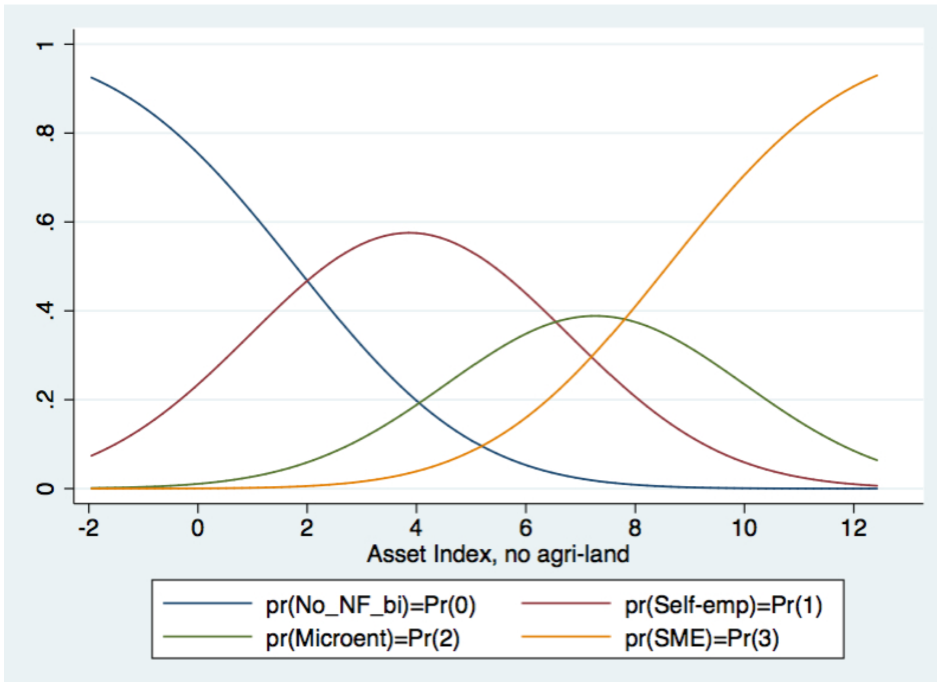


Figure 3.6: Predicted probability of non-farm business types against asset index

Table 3.5: Average marginal effects of order probit estimations of non-farm business types in 2010

Dependent variables	(1) using rainfall shocks				(2) using CV of rainfall			
NF business types:	No NF biz AME/se	NF SE AME/se	Microent AME/se	SME AME/se	No NF biz AME/se	NF SE AME/se	Microent AME/se	SME AME/se
Asset index 05	-0.1039***	0.0809***	0.0192***	0.0039***	-0.1035***	0.0806***	0.0191***	0.0038***
	(0.0148)	(0.0119)	(0.0035)	(0.0013)	(0.0147)	(0.0118)	(0.0034)	(0.0013)
Asset index^2 05	0.0205***	-0.0160***	-0.0038***	-0.0008**	0.0204***	-0.0159***	-0.0038***	-0.0008**
	(0.0053)	(0.0043)	(0.0010)	(0.0003)	(0.0054)	(0.0043)	(0.0010)	(0.0003)
Owned agri land	0.0007***	-0.0006***	-0.0001***	-0.00003**	0.0007***	-0.0006***	-0.0001***	-0.00003**
(100Tarang-wa)	(0.0002)	(0.0002)	(0.00005)	(0.00001)	(0.0002)	(0.0002)	(0.0000)	(0.00001)
age HH head 05	-0.0169***	0.0132***	0.0031***	0.0006**	-0.0171***	0.0133***	0.0032***	0.0006**
	(0.0051)	(0.0040)	(0.0010)	(0.0003)	(0.0051)	(0.0039)	(0.0010)	(0.0003)
age^2	0.0002***	-0.0001***	-0.00003***	-0.00001**	0.0002***	-0.0001***	-0.00003***	-0.00001**
	(0.00005)	(0.00004)	(0.00001)	(0.000002)	(0.0000)	(0.0000)	(0.00001)	(0.000002)
HH members 05	0.0158**	-0.0123**	-0.0029**	-0.0006**	0.0159**	-0.0124***	-0.0029**	-0.0006**
	(0.0062)	(0.0048)	(0.0012)	(0.0003)	(0.0062)	(0.0048)	(0.0012)	(0.0003)
Married 05	-0.0383	0.0298	0.0071	0.0014	-0.0388	0.0302	0.0071	0.0014
	(0.0328)	(0.0257)	(0.0060)	(0.0013)	(0.0329)	(0.0257)	(0.0060)	(0.0013)
HH head: female	-0.0312	0.0243	0.0058	0.0012	-0.0300	0.0234	0.0055	0.0011
	(0.0285)	(0.0222)	(0.0052)	(0.0011)	(0.0282)	(0.0220)	(0.0052)	(0.0011)
HH head Education (base: no schooling)								
educ: primary	-0.0176	0.0137	0.0033	0.0007	-0.0201	0.0156	0.0037	0.0007
	(0.0401)	(0.0312)	(0.0074)	(0.0015)	(0.0403)	(0.0314)	(0.0075)	(0.0015)
educ: secondary	0.0473	-0.0368	-0.0087	-0.0018	0.0426	-0.0332	-0.0079	-0.0016
	(0.0496)	(0.0386)	(0.0093)	(0.0019)	(0.0497)	(0.0387)	(0.0092)	(0.0019)
educ: high school	0.0197	-0.0153	-0.0036	-0.0007	0.0183	-0.0142	-0.0034	-0.0007
	(0.0498)	(0.0388)	(0.0092)	(0.0019)	(0.0500)	(0.0390)	(0.0092)	(0.0019)
educ: college/above	0.1669***	-0.1299***	-0.0308**	-0.0062**	0.1676***	-0.1305***	-0.0309**	-0.0062**
	(0.0630)	(0.0491)	(0.0124)	(0.0030)	(0.0636)	(0.0496)	(0.0124)	(0.0030)
HH farm labor ratio	0.4592***	-0.3574***	-0.0847***	-0.0170***	0.4605***	-0.3585***	-0.0849***	-0.0171***
	(0.0398)	(0.0314)	(0.0140)	(0.0052)	(0.0397)	(0.0313)	(0.0138)	(0.0052)
HH NF worker ratio	0.4278***	-0.3329***	-0.0790***	-0.0159***	0.4282***	-0.3333***	-0.0790***	-0.0159***
	(0.0478)	(0.0384)	(0.0139)	(0.0049)	(0.0475)	(0.0382)	(0.0135)	(0.0049)
HH NF w/o pay ratio	-0.1890**	0.1471**	0.0349**	0.0070*	-0.1837**	0.1430**	0.0339**	0.0068*
	(0.0798)	(0.0630)	(0.0146)	(0.0037)	(0.0800)	(0.0631)	(0.0146)	(0.0037)

Dependent variables		(1) using rainfall shocks				(2) using CV of rainfall		
NF business types:	No NF biz AME/se	NF SE AME/se	Microent AME/se	SME AME/se	No NF biz AME/se	NF SE AME/se	Microent AME/se	SME AME/se
Average farm	0.0342	-0.0266	-0.0063	-0.0013	0.0321	-0.0250	-0.0059	-0.0012
earnings (10k baht)	(0.0326)	(0.0254)	(0.0060)	(0.0013)	(0.0325)	(0.0253)	(0.0060)	(0.0013)
Average NF wage	0.0374**	-0.0291**	-0.0069**	-0.0014*	0.0372**	-0.0290**	-0.0069**	-0.0014*
(10k baht)	(0.0180)	(0.0140)	(0.0035)	(0.0008)	(0.0181)	(0.0140)	(0.0035)	(0.0008)
Interest rate	0.0003	-0.0002	-0.00005	-0.00001	-0.0002	0.0001	0.00003	0.00001
(Provincial avg.)	(0.0013)	(0.0010)	(0.0002)	(0.00005)	(0.0012)	(0.0010)	(0.0002)	(0.00005)
Ratio of HH receiving credits	-0.0960**	0.0747**	0.0177**	0.0036*	-0.1040***	0.0810***	0.0192**	0.0039**
	(0.0397)	(0.0309)	(0.0077)	(0.0018)	(0.0398)	(0.0310)	(0.0078)	(0.0019)
Excess rainfalls (ER)	0.0012**	-0.0010**	-0.0002**	-0.00005*				
	(0.0005)	(0.0004)	(0.0001)	(0.00002)				
Deficit rainfalls	0.0003	-0.0003	-0.0001	-0.00001				
(DR)	(0.0004)	(0.0003)	(0.0001)	(0.00002)				
CV of monthly rains					0.1476*	-0.1149*	-0.0272*	-0.0055*
					(0.0782)	(0.0605)	(0.0153)	(0.0033)
Pseudo R2	0.1854				0.1908			
N. of obs	2048				2048			

***P<0.01, **P<0.05, and *P<0.10. Inverse probability weights to correct attrition bias are applied in the estimation. Variance-covariance matrices are clustered at village level. Other variables include rural characteristics and provincial dummies.

3.5.3 Multinomial logit results of changes in non-farm business status.

In Table 3.6, we explore the relationship of asset/wealth and dynamics of household non-farm business status on startup, exit, and maintaining business between 2006 and 2010. We treat each change in non-farm business status independently by the assumption of the independence of irrelevant alternatives (IIA), an assumption supported by Hausman tests (without clustering standard errors) and Wald tests. The estimated average marginal effects of asset index in panel (1) are positively correlated with entering and maintaining a non-farm business, but negatively correlated with no non-farm business and exiting the sector. The turning point in the response probability for each change in non-farm business status is at different asset index value: 1.92 (minimum), 0.81 (maximum), 0.69 (minimum), 2.65 (maximum) for no non-farm business, entry, exit, and remaining, respectively. It needs more assets for household to maintain their status in operating non-farm business. Households with agricultural land holdings are more likely to have no non-farm business or to leave non-farm business in 2010. A college degree increases the likelihood that a household remains in the no non-farm business status. Households that have previously allocated labor to farming or non-farm employment are more likely to stay the same, having no non-farm business in both 2006 and 2010, indicating occupational persistence. But if households with no non-farm business in 2006 previously experienced working in non-farm wage sector, they have more chance to operate non-farm business in 2010. Higher average wage rates in the non-farm sector increase the chance that households will stop operating a non-farm business and move to work in the wage sector. Households in the community with more credit access are more likely to start a non-farm business, emphasizing the importance of credit in business startup and possibly implying the existence of credit constraint. Exposure to excess rainfall increases the possibility not to start running a non-farm business which could imply that rural non-farm businesses are affected by rainfall shocks. We estimated a model with the cv of rainfall, but the average marginal effects are not statistically significant. Panel (2) uses household wealth instead of asset index. The results are similar to panel (1).

As mentioned earlier, we also estimated bivariate probit models to test for correlation between 2006 and 2010 non-farm business operation. A Wald test rejects the null hypothesis that the error terms of the two binary participation models are uncorrelated. Appendix Table C3 reports estimated average marginal effects for each combination of events. The results are very similar to the multinomial logit estimations in Table 3.6. So we do not elaborate on them here but interpret these only as a robustness check on the multinomial logit results.

Table 3.6: Average marginal effects of multinomial logit estimations of changes in non-farm business types between 2006 and 2010

Dependent variables	(1) Asset index				(2) Wealth			
Changes in NF business types:	No NF biz AME/se	Enter AME/se	Exit AME/se	Remain AME/se	No NF biz AME/se	Enter AME/se	Exit AME/se	Remain AME/se
Asset index 05	-0.0949***	0.0247**	-0.0072	0.0773***				
	(0.0158)	(0.0108)	(0.0078)	(0.0141)				
Asset index^2 05	0.0247***	-0.0153	0.0052***	-0.0146***				
	(0.0075)	(0.0097)	(0.0015)	(0.0047)				
Wealth 05					-0.0773***	-0.0068	-0.0001	0.0842***
					(0.0278)	(0.0230)	(0.0132)	(0.0208)
Wealth^2 05					0.0071**	0.0010	0.0010	-0.0091***
					(0.0033)	(0.0022)	(0.0012)	(0.0032)
Owned agri land	0.0006***	-0.0003	0.0002***	-0.0004**	0.0005**	-0.0002	0.0002***	-0.0005**
(100Tarang-wa)	(0.0002)	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0001)	(0.0002)
age HH head 05	-0.0209***	0.0001	0.0058*	0.0150***	-0.0229***	0.0014	0.0061**	0.0154***
	(0.0058)	(0.0036)	(0.0031)	(0.0047)	(0.0056)	(0.0035)	(0.0030)	(0.0045)
age^2	0.0002***	-0.00001	-0.0001**	-0.0001**	0.0002***	-0.00002	-0.0001**	-0.0001***
	(0.0001)	(0.00003)	(0.00003)	(0.00005)	(0.0001)	(0.00003)	(0.00003)	(0.00004)
HH members 05	0.0137*	0.0060	0.0036	-0.0232***	0.0078	0.0071	0.0038	-0.0187***
	(0.0072)	(0.0052)	(0.0040)	(0.0057)	(0.0073)	(0.0054)	(0.0041)	(0.0058)
Married 05	-0.0701*	-0.0205	0.0010	0.0896***	-0.0802**	-0.0186	0.0005	0.0983***
	(0.0368)	(0.0228)	(0.0209)	(0.0313)	(0.0367)	(0.0226)	(0.0207)	(0.0309)
HH head: female	-0.0479	0.0118	-0.0115	0.0477**	-0.0509*	0.0119	-0.0115	0.0505**
	(0.0297)	(0.0234)	(0.0188)	(0.0213)	(0.0301)	(0.0233)	(0.0190)	(0.0217)
Educ: primary	-0.0358	-0.0105	0.0112	0.0351	-0.0524	-0.0049	0.0090	0.0483
	(0.0428)	(0.0317)	(0.0292)	(0.0311)	(0.0448)	(0.0337)	(0.0291)	(0.0329)
Educ: secondary	0.0040	-0.0101	0.0153	-0.0092	-0.0299	0.0017	0.0127	0.0155
	(0.0527)	(0.0367)	(0.0330)	(0.0396)	(0.0550)	(0.0383)	(0.0328)	(0.0410)
Educ: high school	0.0888	-0.0462	-0.0508	0.0081	0.0454	-0.0336	-0.0538	0.0420
	(0.0559)	(0.0395)	(0.0382)	(0.0398)	(0.0573)	(0.0428)	(0.0369)	(0.0402)
Educ: college/above	0.2232***	-0.0690	-0.0832	-0.0710	0.1976***	-0.0651	-0.0780	-0.0545
	(0.0710)	(0.0649)	(0.0544)	(0.0533)	(0.0730)	(0.0663)	(0.0533)	(0.0587)
HH farm labor ratio	0.5409***	0.0254	-0.1034***	-0.4629***	0.5688***	0.0209	-0.0984***	-0.4913***
	(0.0451)	(0.0298)	(0.0327)	(0.0418)	(0.0459)	(0.0301)	(0.0323)	(0.0430)

Dependent variables	(1) Asset index				(2) Wealth			
Changes in NF business types:	No NF biz AME/se	Enter AME/se	Exit AME/se	Remain AME/se	No NF biz AME/se	Enter AME/se	Exit AME/se	Remain AME/se
HH NF worker ratio	0.5156***	0.0623*	-0.0768**	-0.5011***	0.5179***	0.0615*	-0.0728**	-0.5066***
	(0.0491)	(0.0327)	(0.0316)	(0.0420)	(0.0502)	(0.0332)	(0.0318)	(0.0432)
HH NF w/o pay ratio	-0.7517**	0.2196	0.2117**	0.3203**	-0.8095**	0.2456	0.2152**	0.3487**
	(0.3286)	(0.1503)	(0.0951)	(0.1408)	(0.3332)	(0.1528)	(0.0946)	(0.1419)
Average farm	0.0292	-0.0104	0.0185	-0.0374	0.0264	-0.0067	0.0169	-0.0367
earnings (10k baht)	(0.0329)	(0.0208)	(0.0249)	(0.0310)	(0.0327)	(0.0210)	(0.0253)	(0.0308)
Average NF wage	0.0026	-0.0234	0.0237**	-0.0029	0.0005	-0.0210	0.0249**	-0.0045
(10k baht)	(0.0183)	(0.0144)	(0.0102)	(0.0148)	(0.0191)	(0.0142)	(0.0105)	(0.0159)
Interest rate	-0.0022	0.0011	0.0009	0.0002	-0.0027	0.0011	0.0008	0.0009
(Provincial avg.)	(0.0019)	(0.0013)	(0.0011)	(0.0009)	(0.0019)	(0.0013)	(0.0011)	(0.0009)
Ratio of HH	-0.1047**	0.0526*	0.0274	0.0247	-0.1144***	0.0545*	0.0252	0.0348
receiving credits	(0.0411)	(0.0306)	(0.0279)	(0.0347)	(0.0422)	(0.0314)	(0.0275)	(0.0356)
Excess rainfalls (ER)	0.0012**	-0.0009*	0.0001	-0.0004	0.0010*	-0.0009*	0.0002	-0.0003
	(0.0006)	(0.0005)	(0.0003)	(0.0004)	(0.0006)	(0.0005)	(0.0003)	(0.0004)
Deficit rainfalls	-0.0005	-0.0001	0.0004	0.0003	-0.0006	-0.0002	0.0004	0.0003
(DR)	(0.0004)	(0.0004)	(0.0002)	(0.0004)	(0.0004)	(0.0004)	(0.0002)	(0.0004)
Pseudo R2	0.2348				0.2250			
N. of obs	2048				2048			

***P<0.01, **P<0.05, and *P<0.10. Inverse probability weights to correct attrition bias are applied in the estimation. Variance-covariance matrices are clustered at village level. Other variables include rural characteristics and provincial dummies.

3.6 Conclusions

This study explores the correlates and dynamics of rural non-farm businesses in Thailand. We find strong household wealth effects on the likelihood of running a rural non-farm business. Wage rates are also strongly negatively related to small business creation and maintenance. We also observe more reduction than growth in rural non-farm firm size over the period studied.

Ordered probit model estimations show that both push and pull factors are associated with the likelihood of households operating a non-farm business or hiring more workers in rural Thailand. Non-farm businesses who hire more workers have more household assets, smaller agricultural land holdings, and lower opportunity costs of labor related to other household activities. Besides household constraints on labor allocation, average non-farm wage within community reflects an opportunity cost of household in allocating labors to operating non-farm business. Higher non-farm wage increases the likelihood of households discontinuing non-farm business operation, but households also need higher degree of education, particularly college level, to be able to move to non-farm wage sector. Credit availability also plays an important role in starting business, reflecting borrowing constraint that rural households facing. Moreover, rural non-farm businesses in Thailand are likely to be affected by exposure to rainfall risk. Excess rainfall discourages households to start non-farm business. Hence, rural non-farm business might not be a good decision in diversifying income strategies, compared to working for (skilled) non-farm wage sector. Although the study has a limitation in explaining these factors in casual interpretation, it can give some insights in explaining rural non-farm business state and dynamics in rural Thailand.

REFERENCES

Banerjee, A. and A. Newman (1993). ”Occupational Choice and the Process of Development”. Journal of Political Economy 101: 274-298.

Buera, F. (2009). “A Dynamic Model of Entrepreneurship with Borrowing Constraints: Theory and Evidence”, *Annals of Finance* 5(3): 443-464.

de Mel, S., D. McKenzie, and C. Woodruff (2008). “Who are the Microenterprise Owners?: Evidence from Sri Lanka”, World Bank Working paper 4635.

Fafchamps, M. (1994). “Industrial Structure and Microenterprises in Africa”, *Journal of Developing Areas* 29(1): 1-30.

Gentry, W.M. and R.G. Hubbard (2004). “Entrepreneurship and Household Saving”, *Advances in Economic Analysis and Policy*, 4(1).

Hazell, P. B.R., S. Haggblade, and Thomas Reardon (2007). “Structural Transformation of the Rural Nonfarm Economy”, in S.Haggblade, P.B.R. Hazell, and T. Reardon, eds., *Transforming the Rural Nonfarm Economy* (Baltimore: Johns Hopkins University Press).

Hurst, E. and A. Lusardi (2004). “Liquidity Constraints, Household Wealth, and Entrepreneurship”, *Journal of Political Economy* 112(2): 319-347.

Liedholm, C. (2007). “Enterprise Dynamics in the Rural Nonfarm Economy”, in S.Haggblade, P.B.R. Hazell, and T. Reardon, eds., *Transforming the Rural Nonfarm Economy* (Baltimore: Johns Hopkins University Press).

Karaivanov, A. (2010). “Financial Constraints and Occupational Choice in Thai Villages”, Revised manuscript.

Mead, D.C. and C. Liedholm. (1998). "The Dynamics of Micro and Small Enterprises in Developing Countries", *World Development* 26(1): 61-74.

Miranda, A. and S. Rabe-Hesketh (2006). "Maximum Likelihood Estimation of Endogenous, Switching, and Sample Selection Models for Binary, Ordinal, and Count Variables", *The Stata Journal*, 6(3): 285-308.

Mondragon-Velez, C. and X. Pena-Parga. (2010). "Business Ownership and Self-Employment in Developing Economies: The Colombian Case" in J.Lerner and A. Schoar, eds., *International Differences in Entrepreneurship* (Cambridge, MA: National Bureau for Economic Research).

Nelson, L. K. (2011). "From Loans to Labor: Access to Credit, Entrepreneurship and Child Labor", Working paper.

Quadrini, V. (2000). "Entrepreneurship, Saving, and Social Mobility." *Review of Economic Dynamics* 3: 1-40.

Schoar, A. (2010). "The Divide Between Subsistence and Transformational Entrepreneurship", *Innovation Policy and the Economy* 10(1): 57-81.

Toth, R. (2011). "Credit Misplaced? Testing for Household-level Financial Constraints to Enterprise Activity", NEUDC 2011.

Wang, Shing-Yi. (2010). "Credit Constraints, Job Mobility and Entrepreneurship: Evidence from a Property Reform in China", New York University working paper.

APPENDIX 3A

ATTRITION ANALYSIS

This appendix explains how this study addresses attrition and selection problems in the study. Total rural households surveyed in an initial round compose of 3,600 households. However, after meeting our criteria (having at least one family member employed in all 4 waves), only 2,101 households appear in all four waves. A concern that non-random attrition might bias inferences based on this balanced panel might rise. Moreover, we restrict our sample to households whose members were employed in all four rounds to be consistent with the first paper (avoid conflating transitions between unemployment and employment with transitions among occupations). As a consequence, selection bias could still be an issue if there are significant differences in characteristics between original households and remaining households.

When household characteristics between households in the retained subsample and those who were initially surveyed in 2005 are statistically significantly different, suggesting that the subsample might not be a representative of the original survey. To solve this problem, we follow the procedure suggested by Baulch and Quisumbing (2011), estimating an attrition probit regression, and using inverse probability weights, as suggested by Fitzgerald et al. (1998) and Wooldridge (2002) in adjusting all regressions.

We first run a probit regression where the dependent variable takes the value of zero for a household having at least one member employed and remained in all four waves, and takes the value of one for a rural household registering in 2005 survey, but missing after the first wave or becoming unemployed. Then, we regress that dichotomous dependent variable on baseline variables that could affect the likelihood of attrition. The attrition probit regressions showed in Table A1 include household characteristics, household owned agricultural land and asset index, log earnings in the initial survey round, rural village attrition rate,[29] dummy variables for each of the five main regions in Thailand. The pseudo R^2 statistic can be interpreted as the proportion of attrition that is non-random. Wald tests are then performed to test whether observables jointly explain the predicted attrition probability. As shown in Table A1, the pseudo R^2 values are relatively low. Log household income in 2005 and demographic variables are statistically significantly different from zero at the 1% level, and jointly explain the attrition rate. We can strongly reject the null hypothesis that attrition is unrelated to household characteristics. These tests suggest that non-random attrition could bias the main estimations. We then use the inverse probability weighted method to correct for possible biases.

[29] Village attrition rate is the ratio of total individuals dropped out of the survey after the first wave to total individuals in the village in 2005.

Table A1: Attrition probit regressions

	(1)
Pseudo-R^2	0.2309
Wald Tests	
Rural village attrition rate *(P-value)*	0.000
Log (HH income) 05 *(P-value)*	0.000
Assets *(P-value)*	0.115
Demography and Education[a] *(P-value)*	0.000
Other variables[b] *(P-value)*	0.000
No. of observations	3,600

[a] Demography and education variables include household head's age, age^2, marital status, number of household members, ratio of female members, and dummies for household head's highest educational achievement. [b] Other variables include demographic and education variables, asset variables, and region dummies. Variance-covariance matrices are clustered at sub-district level.

Following Baulch and Quisumbig (2011), we create the ratio of predicted values from the restricted regression and unrestricted regression of reversed attrition probit (RA = 1 if non-attrition, where the unrestricted regression include the same explanatory variables as the attrition probits in Table A1, while the restricted regression excludes the auxiliary variables (demographic variables, household asset variables, rural village attrition rate, and region dummies) in the first period. The inverse probability weights vary from 0.47 to 53.35 with mean 0.88.

Table A2 exhibits probit results of having non-farm business in 2010, comparing the models using a survey weight and an inverse probability weight. The results show similarity on both models with higher pseudo R^2 on the attrition weight models. Although attrition bias indeed appears to exist, it seems that it does not matter much to our estimation results and resulting inferences.

Table A2: Probit estimations of non-farm business participation

	Survey weights b/se	Attrition weights b/se	Survey weights b/se	Attrition weights b/se
Asset index	0.1057***	0.1087***		
	(0.0145)	(0.0148)		
Asset index^2	-0.0256***	-0.0183***		
	(0.0059)	(0.0057)		
Net wealth (million baht)			0.0427**	0.0270*
			(0.0218)	(0.0153)
Net wealth^2			-0.0035	-0.0003
			(0.0028)	(0.0022)
Owned agri land (100rai)	-0.0006**	-0.0005*	-0.0007**	-0.0005**
	(0.0003)	(0.0002)	(0.0003)	(0.0002)
HH head age	0.0158***	0.0084*	0.0198***	0.0118**
	(0.0054)	(0.0050)	(0.0056)	(0.0050)
HH head age^2	-0.0001***	-0.0001	-0.0002***	-0.0001*
	(0.0001)	(0.0000)	(0.0001)	(0.0000)
HH members	-0.0132*	-0.0237***	-0.0073	-0.0188***
	(0.0076)	(0.0069)	(0.0076)	(0.0068)
Married	0.0422	0.0596**	0.0571*	0.0761***
	(0.0313)	(0.0280)	(0.0318)	(0.0289)
HH female ratio	0.1274**	0.1267**	0.1850***	0.1827***
	(0.0528)	(0.0526)	(0.0538)	(0.0536)
HH head Education (Less than primary school/none:base)				
educ: primary	0.0365	0.0225	0.0501	0.0127
	(0.0438)	(0.0331)	(0.0475)	(0.0342)
educ: secondary	-0.0157	-0.0218	0.0237	0.0012
	(0.0544)	(0.0452)	(0.0590)	(0.0468)
educ: high school	-0.0183	-0.0589	0.0419	-0.0144
	(0.0522)	(0.0448)	(0.0565)	(0.0455)
educ: college/above	-0.1227*	-0.1882***	-0.0653	-0.1322**
	(0.0704)	(0.0607)	(0.0744)	(0.0631)
Farm income (10k baht)	-0.0406*	-0.0501**	-0.0211	-0.0271
	(0.0208)	(0.0223)	(0.0205)	(0.0216)
Wage income (10k baht)	-0.0121	-0.0147	-0.0036	-0.0028
	(0.0138)	(0.0142)	(0.0109)	(0.0104)
HH farm labor ratio	-0.4620***	-0.4915***	-0.5203***	-0.5645***
	(0.0429)	(0.0429)	(0.0449)	(0.0452)
HH NF worker ratio	-0.4552***	-0.4897***	-0.4799***	-0.5242***
	(0.0583)	(0.0561)	(0.0572)	(0.0530)
HH NF w/o pay ratio	0.3801***	0.4538***	0.4853***	0.5456***
	(0.1448)	(0.1318)	(0.1490)	(0.1382)
N. of obs	2101	2101	2101	2101
Wald chi2	268.93	310.93	266.90	298.08
Prob > chi2	0.0000	0.0000	0.0000	0.0000
Pseudo R2	0.2204	0.2133	0.1979	0.1904

***P<0.01, **P<0.05, and *P<0.10. Variance-covariance matrices are clustered at village level.

APPENDIX 3B

ASSET INDEX

The asset index used as a proxy for household wealth was constructed using factor analysis based on the method of Sahn and Stifel (2003). Dummy variables for dwelling characteristics and number of durable goods owned are used in the analysis to obtain the first factor. All data come from the SES panel in 2005 (initial year). Table B1 reports summary statistics, factor loadings and scoring coefficients.

Table B1: Asset index summary statistics, factors and scoring coefficients

Variable	Mean	SD	Factor loading	Scoring coef.
Number of rooms	2.73	1.30	0.48	0.09
Housing materials (dummies)				
Cement	0.33	0.47	0.27	0.05
Wood	0.34	0.47	-0.31	-0.06
Others: local/reused materials	0.33	0.47	0.05	0.01
Having electricity	0.99	0.09	0.11	0.02
Cooking fuel (dummies)				
Gas	0.52	0.50	0.53	0.10
Others: Electricity, charcoal, wood	0.45	0.50	-0.52	-0.10
Water supply (dummies)				
Pipe	0.61	0.49	0.09	0.02
Underground water	0.35	0.48	-0.07	-0.01
Others: rain, open sources	0.04	0.19	-0.04	-0.01
Toilet: flush	0.99	0.11	0.11	0.02
Household items (number)				
Microwave	0.08	0.28	0.49	0.09
Refrigerator	0.86	0.47	0.53	0.10
Air condition	0.09	0.40	0.51	0.09
Fan	2.08	1.28	0.61	0.11
Radio	0.71	0.61	0.49	0.09
VCD-DVD player	0.66	0.57	0.56	0.10
Washing machine	0.37	0.50	0.62	0.11
Television	1.13	0.59	0.65	0.12
Cable television	0.01	0.09	0.08	0.01
Satellite dish (for TV)	0.01	0.10	0.21	0.04
Landline	0.17	0.39	0.56	0.10

Variable	Mean	SD	Factor loading	Scoring coef.
Cell phone	0.77	0.85	0.65	0.12
Computer	0.09	0.31	0.62	0.11
Internet access	0.03	0.18	0.48	0.09
Vehicles (number)				
Motorcycle	1.08	0.82	0.36	0.07
Car	0.07	0.28	0.46	0.08
Mini-truck/Van	0.19	0.46	0.45	0.08
2-wheel tractor	0.22	0.44	-0.09	-0.02
4-wheel tractor	0.01	0.11	0.10	0.02
6-wheel or higher	0.01	0.14	0.20	0.04
Owned livestock				
Buffalo	0.04	0.48	-0.07	-0.01
Cow	0.05	0.61	-0.06	-0.01

APPENDIX 3C

ADDITIONAL TABLES

Appendix Table C1: Summary statistics of household's wealth, saving, and debt by changes in business status between 2006 and 2010.

Business status		Wealth (million)		Saving		Debt (million)	
2006 vs. 2010		mean	sd	mean	sd	mean	sd
No NF business	2005	0.44	0.78	1957.18	4435.15	0.14	0.53
	2006	0.42	0.59	2226.74	6781.05	0.13	0.35
	2007	0.43	0.64	2017.27	3852.09	0.13	0.34
	2010	0.51	0.61	2508.29	7056.55	0.16	0.38
Enter	2005	0.44	0.73	1688.78	2616.61	0.10	0.19
	2006	0.44	0.72	1349.00	2569.79	0.10	0.19
	2007	0.43	0.55	2073.41	3618.29	0.11	0.19
	2010	0.49	0.49	3007.87	7485.21	0.21	0.79
Exit	2005	0.60	1.33	2169.96	4814.76	0.20	0.74
	2006	0.63	1.44	1923.86	4025.20	0.12	0.22
	2007	0.47	0.55	2227.55	3989.68	0.19	0.86
	2010	0.59	0.71	2469.82	5837.54	0.14	0.29
Remain	2005	0.69	0.86	2432.15	6799.86	0.21	0.64
	2006	0.72	0.95	2542.54	4219.93	0.17	0.34
	2007	0.81	1.17	2993.92	4724.30	0.18	0.39
	2010	0.83	0.94	3476.09	5617.21	0.20	0.41

Appendix Table C2: Ordered probit (Coefficient reports)

	Model (1) Coef.	Std. err.	Model (2) Coef.	Std. err.	Model (3) Coef.	Std. err.
Asset index 05	0.3844***	0.0570	0.4334***	0.0667	0.3823***	0.0566
Asset index^2 05	-0.0759***	0.0198	-0.0528**	0.0235	-0.0755***	0.0199
Owned agri land (100rai)	-0.0027***	0.0009	-0.0025**	0.0011	-0.0027***	0.0009
age HH head 05	0.0626***	0.0192	0.0640***	0.0192	0.0632***	0.0189
age^2	-0.0006***	0.0002	-0.0006***	0.0002	-0.0006***	0.0002
HH members 05	-0.0584**	0.0229	-0.0635***	0.0226	-0.0587**	0.0229
Married 05	0.1417	0.1215	0.1529	0.1204	0.1432	0.1213
HH head: female	0.1154	0.1051	0.1237	0.1046	0.1110	0.1040
HH head Education 05						
educ: primary	0.0653	0.1481	0.0733	0.1472	0.0741	0.1487
educ: secondary	-0.1750	0.1843	-0.1630	0.1829	-0.1574	0.1842
educ: high school	-0.0729	0.1845	-0.0656	0.1846	-0.0675	0.1849
educ: college/above	-0.6175***	0.2354	-0.5943**	0.2333	-0.6193***	0.2370
HH farm labor ratio 2005	-1.6989***	0.1602	-1.7159***	0.1614	-1.7014***	0.1599
HH NFworker ratio 2005	-1.5827***	0.1908	-1.5971***	0.1908	-1.5819***	0.1893
HH NF w/o pay ratio 2005	0.6993**	0.2956	0.6814**	0.2946	0.6785**	0.2960
Average farm earnings 05 (10k baht)	-0.1264	0.1209	-0.1838	0.1599	-0.1188	0.1204
Average wage 05 (10k baht)	-0.1385**	0.0665	-0.0636	0.1013	-0.1376**	0.0667
Interest rate (Provincial avg.)	-0.0009	0.0047	-0.0052	0.0049	0.0006	0.0045
Ratio of HH receiving credits	0.3551**	0.1466	0.3414**	0.1515	0.3843***	0.1466
Excess rainfall (ER)	-0.0046**	0.0019	-0.0003	0.0044		
Deficit rainfall (DR)	-0.0012	0.0015	-0.0084***	0.0032		
CV of monthly rains during 2005-09					-0.5452*	0.2900
Rural characteristics 05						
- Year-round well usable road	0.0462	0.0781	0.0368	0.0785	0.0531	0.0778
- Ratio concrete of convenient route	0.1040	0.2022	0.0210	0.2020	0.1272	0.1922
- Travel time to nearest city	-0.0033	0.0026	-0.0028	0.0027	-0.0033	0.0026
- Bad soil	-0.1834*	0.0948	-0.2000**	0.0946	-0.1706*	0.0964
- Insuff water for agri.	0.0311	0.0844	0.0126	0.0847	0.0256	0.0841
- Stagnant flood	-0.1186	0.1069	-0.1436	0.1067	-0.1054	0.1070
Region: Central	-0.2703	0.1832	-0.3378*	0.1775	-0.2351	0.1863
Region: North	-0.4536**	0.1959	-0.5231***	0.1949	-0.4010**	0.1996
Region: Northeast	-0.2901	0.1940	-0.3510*	0.1891	-0.2731	0.1979
Region: South	-0.0068	0.1814	-0.1015	0.1772	0.0074	0.1861
Asset index*ER			-0.0027	0.0026		
Asset index^2*ER			-0.0015	0.0014		
Owned agri land*ER			-0.0000	0.0000		
Farm earnings*ER			0.0052	0.0075		
Wage*ER			-0.0052	0.0039		
Asset index*DR			-0.0004	0.0028		
Asset index^2*DR			-0.0004	0.0008		
Owned agri land*DR			0.0001**	0.0000		
Farm earnings*DR			0.0089	0.0063		
Wage*DR			0.0062**	0.0025		
Constant						
Self-employed w/o workers	1.5206**	0.6234	1.5222**	0.6296	1.5814**	0.6220
Microenterprises (<10)	3.1181***	0.6231	3.1300***	0.6262	3.1772***	0.6226
SMEs	4.1343***	0.6306	4.1522***	0.6365	4.1927***	0.6304
Pseudo R2	0.1854		0.1908		0.1842	
N. of obs	2048		2048		2048	

***P<0.01, **P<0.05, and *P<0.10. Variance-covariance matrices are clustered at village level.

Table C3: Bivariate probit estimations (average marginal effects)

	No NF business (P_1=0, P_2=0)		Enter (P_1=0, P_2=1)		Exit (P_1=1, P_2=0)		Remain (P_1=1, P_2=1)	
	AME	Std. err.	AME	Std. err.	AME	Std. err.	AME	Std. err.
Asset index 05	-0.0934***	0.0151	0.0284***	0.0079	0.0001	0.0067	0.0650***	0.0107
Asset index2 05	0.0138***	0.0041	-0.0096***	0.0033	0.0051**	0.0023	-0.0093***	0.0028
Owned agri land (100rai)	0.0004**	0.0002	-0.0003**	0.0001	0.0002*	0.0001	-0.0003**	0.0001
age HH head 05	-0.0235***	0.0052	0.0015	0.0027	0.0054**	0.0023	0.0167***	0.0037
age2_05	0.0002***	0.0001	-0.0000	0.0000	-0.0001**	0.0000	-0.0001***	0.0000
HH members 05	0.0257***	0.0065	-0.0037	0.0038	-0.0039	0.0034	-0.0181***	0.0046
Married 05	-0.0904**	0.0355	0.0009	0.0174	0.0253	0.0168	0.0642**	0.0256
HH head: female	-0.0590**	0.0276	0.0174	0.0162	0.0005	0.0143	0.0411**	0.0194
HH head Education 05								
educ: primary	-0.0609	0.0402	-0.0033	0.0246	0.0207	0.0215	0.0435	0.0284
educ: secondary	-0.0162	0.0492	-0.0124	0.0299	0.0165	0.0256	0.0121	0.0344
educ: high school	0.0204	0.0514	-0.0002	0.0285	-0.0057	0.0243	-0.0145	0.0359
educ: college/above	0.1957***	0.0642	-0.0288	0.0395	-0.0292	0.0335	-0.1377***	0.0447
HH farm labor ratio	0.6181***	0.0390	-0.0329	0.0219	-0.1475***	0.0234	-0.4378***	0.0338
HH NFworker ratio	0.5931***	0.0430	-0.0244	0.0247	-0.1482***	0.0247	-0.4204***	0.0330
HH NF w/o pay ratio	-0.4992***	0.1460	0.0439	0.0653	0.1026	0.0640	0.3527***	0.0998
Average farm earnings (10k baht)	0.0467	0.0316	-0.0197	0.0191	0.0052	0.0191	-0.0322	0.0222
Average NF wage (10k baht)	0.0122	0.0184	-0.0258**	0.0106	0.0210**	0.0088	-0.0073	0.0127
Interest rate (Provincial avg.)	-0.0020	0.0013	0.0004	0.0010	0.0002	0.0009	0.0014	0.0009
Ratio of HH receiving credits	-0.0908**	0.0390	0.0326	0.0229	-0.0047	0.0206	0.0629**	0.0274
CV of monthly rains during 2005-09	0.0558	0.0721	-0.0644	0.0468	0.0450	0.0444	-0.0364	0.0509
Rural characteristics 05								
- Year-round well usable road	-0.0209	0.0217	-0.0023	0.0118	0.0082	0.0108	0.0150	0.0153
- Ratio concrete of convenient route	-0.0005	0.0556	0.0124	0.0304	-0.0116	0.0266	-0.0003	0.0389
- Travel time to nearest city	0.0009*	0.0005	-0.0001	0.0004	-0.0002	0.0004	-0.0007*	0.0004
- Bad soil	0.0341	0.0267	-0.0325**	0.0140	0.0210*	0.0124	-0.0226	0.0186
- Insuff water for agri.	0.0195	0.0222	0.0134	0.0127	-0.0184	0.0113	-0.0146	0.0155
- Stagnant flood	-0.0001	0.0305	-0.0093	0.0159	0.0088	0.0142	0.0005	0.0214
Wald statistics	677.87							
No. of obs	2048							

***P<0.01, **P<0.05, and *P<0.10. Inverse probability weight to correct attrition bias has been applied in the estimation. Variance-covariance matrices are clustered at village level. Other variables include provincial dummies.

CHAPTER 4

INTERSECTORAL AND GENDER HETEROGENEITY IN THE MARGINAL EARNINGS GAINS ASSOCIATED WITH EDUCATION IN INDONESIA

4.1 Introduction

The role of education in sector-specific and overall economic growth is a longstanding theme in development economics. According to Schultz (1964), human capital, in the form of education of rural workers and especially farm managers, is important for agricultural productivity as it facilitates technological change. Lewis (1954) saw the nonfarm sector as potentially dynamic and expanding. In his view, and in subsequent dual economy models that build on his insights, excess supply of farm labor with low marginal returns provides a surplus labor supply available to employers in the industrial sector who pay a higher wage because labor is more productive when mixed with industrial capital. In the course of structural transformation, higher labor productivity in agriculture allows households to allocate more time to non-farm activities or to migrate to the urban industrial sector (Timmer, 2009). Anticipating this, rural households invest their profits from agriculture in human capital formation, education in particular, also raising non-farm labor productivity. All these mechanisms lead to a growing rural nonfarm sector and gradual shrinkage in the intersectoral differences in the marginal returns to labor (Barrett et al., 2010).

Investment in human capital affects labor productivity and therefore earnings. In human capital theory, the optimal human capital investment is at the point where the expected present value of marginal gain in private returns equals the present value of marginal opportunity cost. As described by Card (1999), studies of the role education plays in determining earnings almost always employ the human capital earnings function formulation of Mincer (1974). While the standard Mincerian model is often applied in a way that implicitly assumes that returns to schooling are constant across grade years and for all persons (conditional on other covariates), recent studies on the estimated returns to schooling allow returns to vary systematically across individuals due to individual and family characteristics, as well as to skills that are innate or that develop during early life (Card, 2001; Carneiro, Heckman and Vytlacil, 2001; Carnerio, Heckman, and Vytlacil, 2011; Carneiro et al., 2011).

In low- and middle-income countries additional sources of heterogeneity in the marginal gains from education may arise because of intersectoral differences in the marginal returns to labor. If the marginal product of human capital is low in farm production but relatively high in nonfarm wage work, then the agricultural household will typically increase the share of its labor stock allocated to nonfarm work (Huffman and Orazem 2007). But household ownership of sector-specific fixed factors of production, like farmland or farming skills, combined with spatially heterogeneous access to public schools, transportation and communications infrastructure that affect labor productivity could impede rural individuals' ability and willingness to invest in education and to reallocate labor intersectorally. The result can be persistent differences both in the likelihood of adult employment in different sectors and in the earnings gains from education across sectors. These joint differences remain unexplained in the literature, however. This paper, therefore, explores how individual earnings respond to added educational attainment given intersectoral labor market heterogeneity and gender differences, as well as the potential signaling that comes from educational credentialing in the job market.

Estimation of those prospective intersectoral differences in the earnings gains from education is complicated by two empirical regularities of developing country labor markets. First, educational attainment plays a key role in sorting on the job market, by influencing employability in different sectors. For example, farmers and lawyers have markedly different educational requirements. In societies where many people do not complete primary or secondary school, a large share of the workforce effectively lacks access to salaried or waged jobs in the formal private and public sectors by virtue of their limited education and must rely on self-employment in the agricultural or non-farm sectors. Under such circumstances the marginal impact of education on the probability of formal sector employment may be as or more important than the effect of education on earnings within any given sector. Put differently, education may play a role both in sorting workers among sectors and in determining their earnings within a sector.

Second, rates of self-employment are especially high in developing countries. At least since Spence (1973) it has been well known that in the presence of asymmetric information about individuals' imperfectly observable ability, individuals invest in education not merely to increase their marginal productivity and thereby their future earnings, but also to signal ability or other unobservable characteristics that are exogenous to schooling and that employers might reward (Riley 1979). Especially if education matters to the probability of employment in sectors that reward labor more in economies with imperfect intersectoral labor market integration, then the signaling function of schooling may play a disproportionately large role in determining the correlation between earnings and education. Since the need for signaling arises due to asymmetric information between an individual and her prospective employer, signaling effects are necessarily absent among the self-employed. Perhaps we can and should therefore distinguish between the earnings gains from education that is not due to signaling by differentiating between the gains from education among the self-employed and among the employed.

In this paper we combine these three ideas: education affects the probability of sorting into different employment sectors, it may be associated with different marginal earnings gains in each sector, and the differences in marginal earnings gains may reflect in part gains due to signaling associated with educational attainment, as distinct from any true productivity gain from education. Non-farm employment is generally a more remunerative sector than farming and education is required to secure stable high wage employment. However, higher education is an investment that needs funding and time, as well as foregone earnings during school years. If parents who expect their children to take over the family farm perceive that the earnings gains from education in farming are less than in the non-farm sector, and they understand that the child will not need education as a signal of ability if she is to be self-employed, then they are likely to invest less in schooling for their children than would parents who expect their children to pursue a non-farm career as a salaried civil servant or firm employee, other things equal. In equilibrium, heterogeneous educational attainment might be an equilibrium outcome for individuals whose parents have different expectations regarding their likely ultimate sector of employment and between employed or self-employed status.

In this study, we focus on exploring variation in the Mincerian, private earnings gains associated with education, differentiating between the farm and non-farm sectors of employment and between employment and self-employment (i.e., farm self-employment, farm worker, non-farm self-employment, and non-farm employee). This approach does not estimate the true returns to education in so far as it necessarily omits the costs of education under the assumption that the cost of education are only the earnings foregone while in school or that the direct costs and unmeasured earnings while in school cancel out each other (Fields, 1980). Under these assumptions, the estimated coefficients on educational attainment in the earnings function can be interpreted as private gross rates of return, as per the education and labor literatures, or more precisely as private gross gains from education. However, these estimates must be carefully interpreted as we focus on descriptive explanation of earnings differentials as differences of coefficients on the schooling variables across the four employment sectors, allowing as well for variation by gender in those differences.

An obvious challenge in estimating the Mincerian returns to education disaggregated by sector of employment is that sector and earnings are jointly determined by educational attainment. We therefore need a credible strategy to instrument for sector of employment. Because parents provide references, job information, and social connections that benefit children in the labor market (Magruder, 2010), we use parental information on occupation and education, as instruments influencing children’s occupation selection. We then use the resulting predictions of sector of employment to estimate the marginal earnings gains associated with education conditional on sector of employment.

We find evidence of both sheepskin effects associated with completion of multi-year stages of schooling and, most importantly, pronounced inter-sectoral differences in the marginal earnings gains from schooling. This obviously has significant implications for parents' optimal educational investments in their children if not all children are equally likely to enter all careers. However, within the nonfarm or farm sectors, only those workers who graduate at higher grades of schooling earn more than the self-employed. We furthermore find that this reflects the signaling effect of secondary and tertiary education in non-farm employment Moreover, females with a high school or college degree benefit more from their schooling as their earnings increase substantially when they graduate at least from high school.

The remainder of the paper is organized as follows. Section 2 describes the data and Indonesian educational context. Section 3 describes the econometric model and estimation strategy. Section 4 reports empirical findings on both the occupational attainment and the earnings equations. Section 5 concludes.

4.2 Data and descriptive analyses

The data used in this study are from the Indonesia Family Life Survey (IFLS). The IFLS is a household and community level panel survey, conducted over four rounds in 1993-94, 1997, 2000, and 2007-08. The original round in 1993-94 (IFLS1) surveyed 7,224 households with 33,081 individuals in 13 provinces, representing more than 80% of Indonesians. The original households were re-sampled in the subsequent rounds and the split-offs from the original households were tracked and included in the survey. Hence, the number of individual records increases every round. By the 2007-08 round (IFLS4), the survey covers 13,536 households with 50,579 individuals. Overall, 87.6 percent of households participating in first IFLS wave were interviewed in all four rounds (Strauss et al., 2009). However, there are still many missing values on many variables. We look carefully at attrition issues in Appendix 4A.

Sectors of employment are classified based on respondents' reported primary job. The sample used consists of individuals aged 15-70 years old with self-reported earnings and schooling information, excluding unpaid family workers. The survey asked salary/wage (including the value of all benefits) and net profit (i.e., gross revenues less all business expenses) that wage workers and self-employed individuals received during last month and last year, respectively.[30] No wage is reported for unpaid family workers. Hence, self-employed earnings refer to earnings that result from entrepreneurial effort regardless of having family or non-family workers. We only use earnings from individuals' primary job, not including other sources of income. Years of schooling is defined as the highest grade completed. Since there are some conflicting reports of years of schooling across survey waves, for those age 25 years or older in IFLS1 with no reports of attending school in subsequent survey rounds, we use their schooling reports from IFLS1. We assume that the earliest replies are the most accurate recall information for the older people given that they reported no further schooling thereafter.

[30] We use yearly earnings in our analyses as yearly earnings have fewer negative values than monthly earnings and obviate seasonality issues.

Information on parental education and occupational history, and agricultural land are extracted from IFLS1-2 (1993 and 1997) and then merged with individual earnings records from IFLS4 (2007-8). We thus have data (at least) ten years apart for determinants of ultimate sector of adult children's occupation and current earnings in IFLS4. Parental information on occupation and highest grade of schooling comes from three sources. First, data on parents who were living in the same household in IFLS4 were tracked back from IFLS1-2 self-reports. Second, information on parents non-coresident with the adult child in 2007 is only used for deceased parents. We do not include the group of parents who were not in the same household but still alive since their reports of occupation refer to their current occupation which might be endogenous to the sector of work of the grown child. Third, if there is no information whether their parents lived in the same household or not in IFLS4, or if their parents were still alive but did not live in the same household, we trace back to their relationship reports as a son/daughter of the household head and wife/husband and use the IFLS1 information of household head and wife/husband as their parents' information. Agricultural land value (in 2007) is also drawn from the earlier of IFLS1 or IFLS2 depending on the earliest information we have.

Table 4.1 shows summary statistics for the subsamples from IFLS1 and IFLS2 that we use and the IFLS4 sample. Sample 1, the full sample, contains non-missing values of schooling, occupation, and earnings variables for 15-70 year-old individuals, totaling 12,675 observations. Sample 2, the estimation sample, includes only households with non-missing values for all variables listed in the table, after matching all parents' information, totaling 6,486 individuals. Balancing tests on key variables between each of samples 1 and 2 show mostly statistically significant differences. The missing values do not appear randomly. Appendix 4A explores attrition using inverse probability weights to correct for sample attrition from merging the IFLS4 data with parental data and value of agricultural land data from the earlier survey rounds. Although our results indicate that attrition is likely non-random, the OLS estimates of the earnings equations when using survey weights and attrition weights are statistically insignificantly different. Nonetheless, we use the attrition weights in all of our estimations.

Table 4.2 shows the distribution of sample observations across years of schooling completed, sectors of employment and gender as of 2007. In Indonesia, the elementary level includes grades 1-6, junior high school includes grades 7-9, senior high school includes grades 10-12, and years 13 and higher are university or college level. A majority of individuals complete no more than an elementary school education. These tables show clearly that the probability of working in farming, especially farm self-employment moves inversely with educational attainment while the probability of working in the non-farm employee sector increases with education, for both genders. This relation is as expected since the non-farm employee sector involves more professional training. Females with less than an elementary school education, however, tend to work in non-farm self-employment more than male elementary school dropouts do. Men comprise 63% of the actively explained sample, earn more than women, on average, in each sector, and are more likely to have completed at least junior secondary school. On average, nonfarm workers earn the most per year, followed by non-farm and farm self-employment, while farm workers earn the least.

Table 4.1: Summary statistics of each sample used in the study

	Sample 1: Full sample		Sample 2: Estimation sample	
	Mean	Std. Dev.	Mean	Std. Dev.
Years of schooling	7.49	4.63	7.30**	4.60
log earnings 2007	15.19	1.36	15.11**	1.35
Age	41.51	13.03	41.99**	14.18
Female	0.36	0.48	0.38**	0.49
HH head	0.58	0.49	0.53***	0.50
Dmomse			0.24	0.43
Dmomwkr			0.17	0.38
Ddadse			0.48	0.50
Ddadwkr			0.38	0.48
Mom elementary			0.26	0.44
Mom Jr.high sch			0.04	0.20
Mom Sr. high sch			0.03	0.17
Mom University			0.01	0.08
Dad elementary			0.33	0.47
Dad Jr.high sch			0.06	0.23
Dad Sr. high sch			0.06	0.23
Dad University			0.02	0.12
Value of agri land			22.24	102.40
Durban_07	0.46	0.50	0.46	0.50
No of obs	12,675		6,486	

Survey weights are applied. ***P<0.01, **P<0.05, and *P<0.10 for tests of the equal means null hypothesis between samples 1 and 2. Sample 1 contains non-missing values of non-missing values of schooling, occupation, and earnings variables for 15-70 year-old individuals, totaling 12,675 observations. Sample 2 contains non-missing values of all variables on the list, including parental information.

Table 4.2: Years of Schooling by sectors of employment and gender, 2007

	Male							
Years	Farm worker		Farm SE		NF worker		NF SE	
	N	P	N	P	N	P	N	P
0	54	0.06	210	0.22	65	0.07	65	0.07
1	33	0.11	76	0.26	27	0.09	36	0.12
2	42	0.08	143	0.26	76	0.14	78	0.14
3	33	0.06	142	0.27	77	0.15	76	0.14
4	41	0.07	124	0.22	97	0.17	86	0.15
5	38	0.09	103	0.24	63	0.14	55	0.13
Primary 6	**177**	**0.07**	**482**	**0.19**	**599**	**0.23**	**363**	**0.14**
7	20	0.08	35	0.14	76	0.31	39	0.16
8	10	0.04	41	0.16	93	0.36	54	0.21
Jr. high 9	**93**	**0.06**	**186**	**0.12**	**503**	**0.32**	**270**	**0.17**
10	4	0.03	10	0.08	54	0.41	26	0.20
11	5	0.04	20	0.14	55	0.39	24	0.17
Sr. high 12	**89**	**0.03**	**176**	**0.06**	**1,327**	**0.45**	**435**	**0.15**
13		0.00	9	0.10	35	0.41	12	0.14
14	1	0.01	1	0.01	36	0.47	16	0.21
15	1	0.01	1	0.01	38	0.42	17	0.19
University 16	**6**	**0.00**	**21**	**0.02**	**540**	**0.43**	**131**	**0.11**
Total	647	0.05	1780	0.14	3761	0.30	1783	0.14
log earnings	14.9191		14.9384		15.6583		15.6483	
std. dev.	1.1954		1.2070		1.2940		1.2005	
CV	0.0801		0.0808		0.0826		0.0767	

	Female							
Years	Farm worker		Farm SE		NF worker		NF SE	
	N	P	N	P	N	P	N	P
0	106	0.11	106	0.11	101	0.11	227	0.24
1	21	0.07	27	0.09	27	0.09	51	0.17
2	33	0.06	39	0.07	50	0.09	94	0.17
3	25	0.05	34	0.06	48	0.09	90	0.17
4	35	0.06	17	0.03	58	0.10	101	0.18
5	26	0.06	27	0.06	37	0.08	87	0.20
Primary 6	**95**	**0.04**	**83**	**0.03**	**332**	**0.13**	**418**	**0.16**
7	6	0.02	1	0.00	30	0.12	37	0.15
8	6	0.02	5	0.02	19	0.07	31	0.12
Jr. high 9	**24**	**0.02**	**32**	**0.02**	**265**	**0.17**	**218**	**0.14**
10	2	0.02	0	0.00	17	0.13	18	0.14
11			2	0.01	20	0.14	14	0.10
Sr. high 12	**17**	**0.01**	**25**	**0.01**	**603**	**0.20**	**286**	**0.10**
13	1	0.01	0	0.00	23	0.27	6	0.07
14	0	0.00	1	0.01	17	0.22	5	0.06
15	0	0.00	0	0.00	24	0.26	10	0.11
University 16	**8**	**0.01**	**2**	**0.00**	**468**	**0.38**	**66**	**0.05**
Total	405	0.03	401	0.03	2139	0.17	1759	0.14
log earnings	14.1184		14.1456		15.2390		14.9064	
std. dev.	1.2869		1.3039		1.3921		1.3027	
CV	0.0912		0.0922		0.0914		0.0874	

N = number of observations, P = proportion in sector

4.2.1 Indonesia education background

Between 1973-1974 and 1978-1979, more than 60,000 new primary schools were constructed under the INPRES program. The program targeted children who had not previously been enrolled in school (Duflo, 2001). Children born after 1962 therefore enjoyed at least partial exposure and benefited from increased access to primary schools in Indonesia. The enrollment rate in primary school indeed increased from 72 percent in 1975 to nearly universal coverage by 1995. The enrollment rate in junior high school rose from 18 percent in the 1970s to approximately 62 percent in 2005, while senior high school enrollment rate increased to around 40 percent by 2005 (Granado et al., 2007). A higher enrollment rate in primary school reduces the enrollment gap across income groups. However, inequalities remain at the junior and senior high school levels. Suryadarma and Suyrahadi (2010) also find that children from poor households are less likely to graduate from junior secondary school. Moreover, regional differences are strongly associated with the enrollment gaps in income levels as the richest quintile in Papua still has lower enrollment rates than the poorest quintile in Sumatra (Granado et al., 2007)

4.2.2 Differences in unconditional earnings distributions across employment sectors

Work in the non-farm sector, either in self-employment or as an employee, is significantly more remunerative, on average, than farming. The top panel of Figure 4.1 shows the cumulative distributions of log earnings in 2007 by sector. The farm worker earnings distribution is first-order stochastically dominated by both non-farm self-employment and wage employment, while farm self-employment and both non-farm occupations are not statistically significantly different in dominance ordering, per the Davidson and Duclos (2000) test. Since earnings differences are negligible up through completion of junior high school, we combine all these who did not complete below senior high school in these simple descriptives. A plot of the cumulative earnings distributions for each completion level clearly shows that the earnings distributions of college graduates first-order stochastically dominate the earnings distributions of lower schooling level graduates. But only a small number of individuals graduated from college, we combine all those individuals who completed at least senior high school. As referenced in the bottom panel of Figure 4.1, nonfarm workers' log earnings distribution statistically significantly first-order stochastically dominates farm workers' log earnings distribution among this better-educated subpopulation. For high school and university graduates, we can clearly distinguish the earnings distributions of workers between the nonfarm and farm sectors, reflecting the intersectoral differences in labor returns that underpin dual economy models and other lines of the development literature.

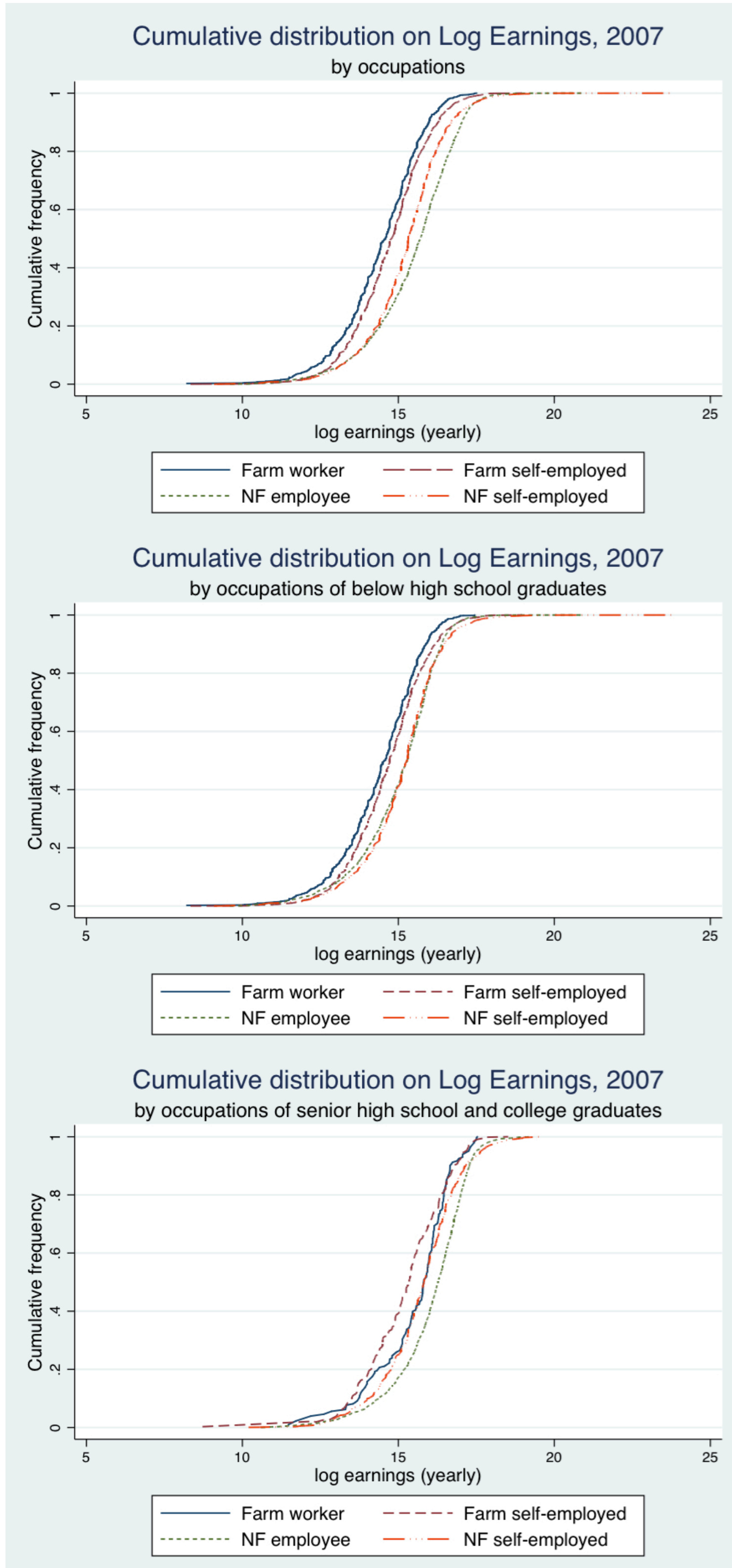


Figure 4.1: Cumulative distributions of log earnings across employment sectors

4.2.3 Differences in education-conditional earnings distributions within each sector

When estimating kernel densities (not shown here) for each sector of employment, all of log earnings distributions shift rightward with successively higher levels of education. These results confirm that higher education is associated with increased earnings irrespective of sector. But these shifts vary by sector of employment. Figure 4.2 depicts log earnings distributions by education level for each sector of employment. The conditional earnings distribution for any education level up to senior high school is statistically first-order stochastically dominated by the conditional earnings distribution for those with university level education in each sector of employment. The nonfarm employee sector (bottom left panel) clearly shows the greatest difference in education-conditional earnings, exhibiting the first-order stochastic dominance of earnings distribution at the university and senior high school levels relative to lower levels of educational attainment. In both the farm and non-farm sectors, the shift in earnings distributions associated with increased educational attainment is far greater among the employed than the self-employed, consistent with the hypothesis that signaling effects play an important role in labor markets.

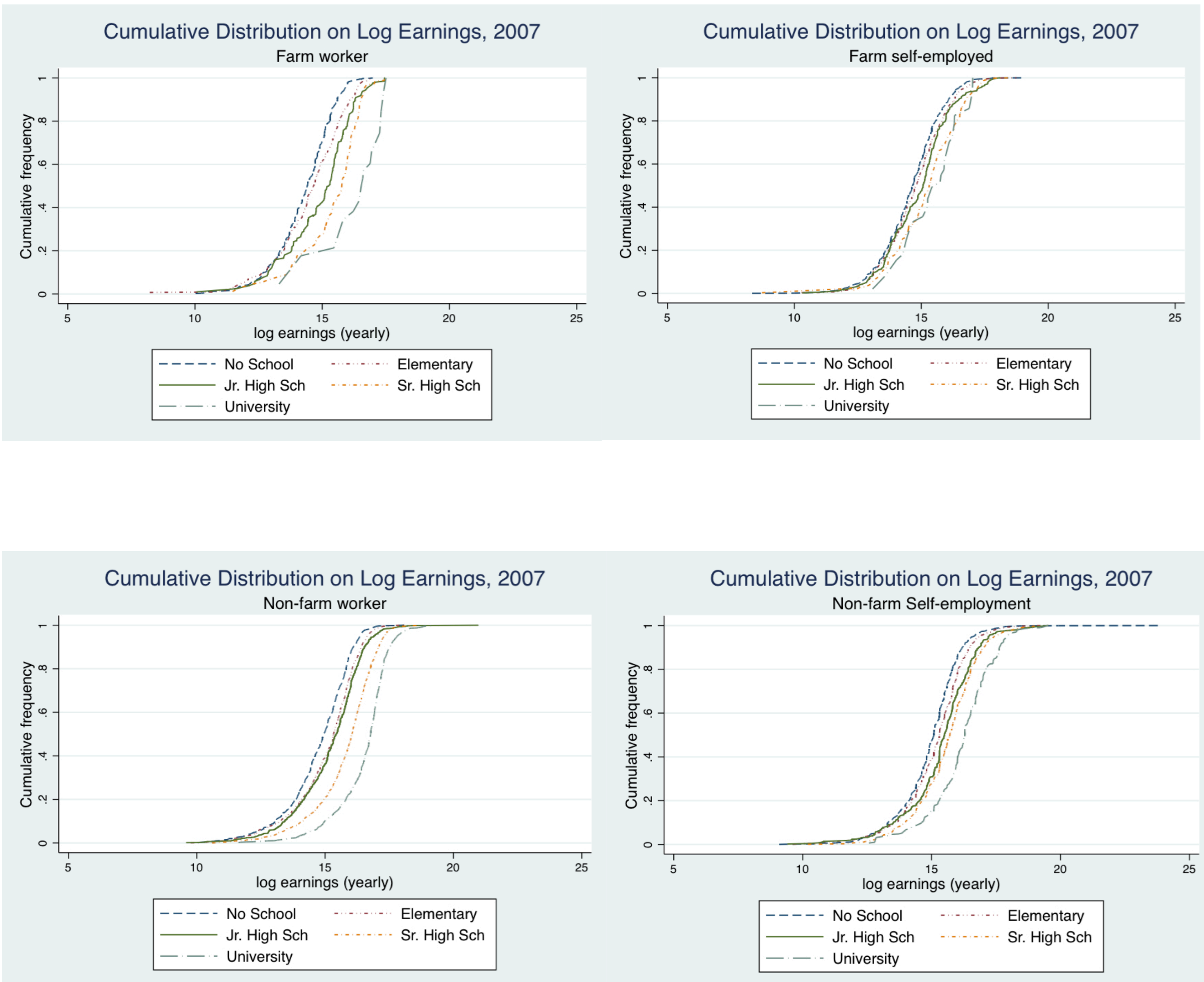


Figure 4.2: Cumulative distributions of education- conditional log earnings, by sector

4.3 Econometric strategy

Ultimately, we want to estimate a Mincerian earnings equation to explore the (private) marginal earnings gains from educational attainment, allowing for variation between the farm and non-farm sectors so as to accommodate structural differences typical of developing economies, between self-employment and wage employment so as to distinguish between the marginal earnings gains and the signaling effects of education on earnings, and between men and women, especially in a predominantly Muslim society where gender may affect labor market access. The econometric challenge is that including employment sectors in the Mincerian earnings equation raises concerns about bias due to both the prospective endogeneity of schooling choices and to selection effects into sector of employment.

We worry less about the endogeneity of schooling choices. In Duflo (2001), OLS estimates of the returns to schooling are 7.8 percent per year in Indonesia while two-stage least squares estimates, using policy change in access to primary schooling as instruments, indicate returns of 7-10 percent. Duflo's results do not support the view that OLS estimates of returns to education in developing countries are biased upward by endogenous schooling choices. Oyelere (2010), using the exogenous timing of policy change in free primary education across regions in Nigeria, also shows estimated returns to education similar in both OLS and IV estimates. Given that endogenous schooling choices might not be a significant source of bias, we then focus on selection into sector of employment.

OLS estimates of the earnings equation could be biased if selection into employment sectors were correlated with regressors in the earnings equation.

Siphambe (2000) and Lassibille and Tan (2005) emphasize the endogenous choice of sector of employment and therefore correct for selection effects using an Inverse Mills ratio correction from the multinomial logit estimator adapted by Lee (1984) from Heckman (1979). Here we use a more efficient correction following Dubin and McFadden (1984) and modified by Bourguignon et al. (2007). Comola and Mello (2013), using Indonesian labor market survey data (Sakernas), apply a multinomial selection for three labor market statuses (inactivity, wage earner, and nonsalaried work) and instrument educational attainment in both the earnings and selection equations. They estimate returns to education of 9.0-10.8 percent depending on whether one applies a multinomial selection procedure or treats schooling years as endogenous.

Magruder (2010) investigates intergenerational networks between parents and their children and finds that network connections of fathers increase employment rates of their sons. Hellerstein and Morrill (2010) show that about one-third of sons and one-fifth of daughters work in the same occupation as their father. Kramarz and Skans (2011) use Swedish employer-employee data to show that family networks, not their classmates, play an important role in searing for boys their first job, particularly in areas that people have low educational attainment on average. The underlying idea in these studies is related to intergenerational mobility (Becker and Tomes, 1979; Black and Devereux, 2010). Parents' race, ability, education, occupation, and other characteristics influence their children's adult work environment. Many studies therefore use parental education as an instrument for child education. We use parents' occupation and education to instrument for their adult child's sector of employment in 2007 in the first stage estimation.

Given these concerns, we explore the Mincerian returns to education following two distinct estimation strategies. First, we estimate by ordinary least squares (OLS), under the naïve assumption that sectors of employment are exogenous, with interactions among sectors, gender, and education in order to test the heterogeneity of marginal earnings gains hypotheses. Second, we treat sorting into each employment sector as a selection problem. We then account for the sector-conditional probability of multinomial selection estimated in the first stage, using parental variables as instruments for sector selection. We then estimate the earnings equations separately for each sector, correcting for the selection terms following Dubin and McFadden (1984) and Bourguignon et al. (2007). We caution, however, that although the sector selection is instrumented, the estimated marginal earnings gains from schooling cannot be interpreted as causal estimates of the effects of education on earnings because the schooling variable is still potentially endogenous. But, mentioned previously, several prior studies show that the estimated returns to schooling are very similar for both OLS and IV estimations, suggesting that such bias is likely small and that these selection-corrected estimates are likely close to the true causal parameter estimates.

4.3.1 Multinomial occupational attainment estimations

Parents provide their children with references, job information, and social connections that benefit children in the labor market. In this study, we use dummy variables for father's and mother's working status as a worker or self-employed to capture the influence of parents' occupation on child's sector of employment. Unfortunately, the information for dead parents only reports working status, with no information on farm or nonfarm sector, so we cannot separate parents by sector of employment. Parental education is defined as dummy variables for completed elementary, junior high school, senior high school, and university, with less than completion of elementary school as a base. We expect that parents with higher educational attainment are likely to work in the nonfarm sector, implying that parental education here could identify sector of employment for their children. Individual's schooling is assumed to be exogenous as we want to explore how education is related to sorting into each sector. Based on the multinomial logit model, we define

$$\Pr(O = j) = \frac{\exp(X\Gamma(j))}{1 + \sum_{j=2}^{4} \exp(X\Gamma(j))}$$

where
$$\begin{aligned} X\Gamma(j)_i = {} & \gamma_0 + \gamma_1 S_{it} + C_i'\gamma_2 + \gamma_3 female + C' \cdot female \cdot \gamma_4 + \gamma_5 age_i + \gamma_6 age_i^2 \\ & + D_i'\gamma_7 + \gamma_8 e_i^k + X_i'\delta + \nu_i \end{aligned} \tag{4.1}$$

To ensure model identification, $\Gamma(j)_i$ is set to zero when the individual job is farm worker (a base case). Then, O is occupational choice in 2007 and j = 2, 3, 4 refer to farm self-employment, non-farm worker, and non-farm self-employment, respectively. S_i is the highest year of school completion and C_i is a set of dummies for degree completion. D_i is a vector of dummy variables of father's and mother's occupation as worker or self-employed, and for parents' highest schooling level. e^k is a sector-specific capital endowment (family agricultural land holdings) and X_i is a vector of other controls (used in the second stage), including household head, urban area, and provincial dummies.

4.3.2 Earnings equation estimations

We estimate the education-earnings relationship and differences across sectors of employment and gender. We follow two distinct estimation strategies, as described above. First, we assume that sectors of employment are exogenous. Second, allowing for the prospective endogeneity of sector of employment, we estimate a selection model of employment sectors.

4.3.2.1 Treating sector of employment as exogenous

If we assume that sector of employment is exogenous, we estimate the earnings equation including interactions between years of schooling, sectors of employment, as well as gender, so as to explore the differences in (private) returns to education across sectors and genders. We also add degree effects or 'sheepskin effects' – i.e., a dummy variables for highest degree completed (Hungerford and Solon, 1987; Belman and Heywood, 1991) – as we see further rightward shifts of earning distributions in each education degree level. The baseline model is defined as follows:

$$\ln y_i = \alpha_i + \beta_1 S_i + \sum_{j=1}^{4}\phi_j C_{ji} + \beta_2 S_i female_i + \sum_{j=1}^{4}\phi_j^f C_{ji} female_i$$
$$+ \sum_{k=1}^{3}\gamma_k O_{ki} + \sum_{k=1}^{3}\gamma_k^f O_{ki}\cdot female_i + \sum_{k=1}^{3}\sum_{j=1}^{4}\gamma_{jk} O_{ki} C_{ji} + \sum_{k=1}^{3}\sum_{j=1}^{4}\gamma_{jk}^f O_{ki} C_{ji}\cdot female_i$$
$$+ \sum_{k=1}^{3}\gamma_k^s O_{ki} S_i + \sum_{k=1}^{3}\gamma_k^{fs} O_{ki} S_i\cdot female_i + \beta_3 age_i + \beta_4 age_i^2 + X_i'\delta + \varepsilon_i \tag{4.2}$$

where ln y_i is natural log of earnings of individual i; S_i is completed years of schooling; *age* is individual age (instead of years of potential experience and the quadratic term for capturing the life cycle effect); O_{ki} is a dummy for employment sector where k = 1,2,3 for farm self-employment, nonfarm employee, and nonfarm self-employment, respectively. Then, all interaction terms are added for both intercepts and slopes, interpreting γ^s and γ^{fs} as additional earnings gains associated with schooling from the base case of male farm worker and female farm worker. C_{ji} is a dummy for degree completion to reflect sheepskin effect in a cumulative way, so C_{1i} = 1 for those who completed grade 6 (primary school graduates), C_{2i} = 1 for grade 9 graduates (junior high school graduates), C_{3i} = 1 for grade 12 graduates (senior high school graduates), and C_{4i} = 1 for university graduates. We also interact degree effect terms with sectors and gender. X_i is a matrix of control variables, including household head, household agricultural land value, urban area, and provincial dummies. Standard errors are clustered at community level and attrition weights are applied.[31]

[31] Given the findings in Duflo (2001), we also introduced dummy variables to split cohorts into three groups: i) cohort1 is a young generation, born after 1972 (indirect exposure to the INPRES program); ii) cohort2 is a generation aged 2-12 in 1974 (direct exposure to the INPRES program); and iii) cohort3 is an older generation, aged 13+ in 1974 (no exposure to the program). These cohort variables shows positive correlations between log earnings and older cohorts, representing age cohort effects rather than the result of more access to primary

4.3.2.3 Selection-corrected estimates of the marginal earnings associated with education

We can also use the selection correction methodology developed by Dubin and McFadden (1984) and modified by Bourguignon et al. (2007). The method seeks to estimate the earnings function for each sector separately, but correcting for prospective selection bias based on multinomial logit instead of the usual Heckman selection model based on a probit estimator. Here, let log earnings in the sth alternative is given by

$$\ln y_s = \alpha_s + Z_s\beta_s + u_s \tag{4.3}$$

where Z contains the main covariates as in equation (4.2), except sector variables and their interactions. Let ln y_s be observed only if alternative s – occupational choice s – is chosen among three alternatives. Following the Heckman selection model, Dubin and McFadden (1984) and Bourguignon et al. (2007) include multiple correction terms to control for self-selection into the sth alternative instead of only an inverse Mills ratio term for self-selection correction. Hence, equation (4.3) becomes

$$\ln y_s = \alpha_s + Z_s\beta_s + h(P_0,...,P_3) + e_s\ , \tag{4.4}$$

where P_j is the probability that alternative j, j = 1, 2, 3, 4 will be chosen. P_j follows the multinomial logit model in equation (4.1).

---

schooling from the INPRES program. We therefore omit these structural breaks in the reported estimates.

Dubin and McFadden (1984) assume that $E(u_s) = \sigma_s \frac{\sqrt{6}}{\pi} \sum_{j \neq s} r_j \left[\eta_j - E(\eta_j) \right]$, where η_j is a disturbance term from random utility maximization, conditional on the alternative s being chosen. Hence, (4.1) becomes

$$\ln y_s = \alpha_s + Z_s \beta_s + \sigma_s \frac{\sqrt{6}}{\pi} \left[\sum_{j \neq s} r_j \left(\frac{P_j \ln(P_j)}{1 - P_j} \right) + r_s \ln(P_s) \right] + e_s \qquad (4.5)$$

where r_j is the correlation coefficient between disturbances u_s and η_j, and e_s is a residual with asymptotic mean zero. Equation (4.5) is estimated in two steps. We first estimate the multinomial logit model for each of occupational choice in (4.1) and then the predicted probabilities are substituted into the selection correction terms in equation (4.5). Standard errors are bootstrapped with 1000 replications to account for the generated regressors.

4.4 Empirical findings

Table 4.3 reports the estimated average marginal effects from the multinomial logit in equation (1). Parents' occupation and education indeed influence the employment status of individuals. These variables statistically significantly jointly determine the estimated probability of an adult child's sector of employment. Individuals whose father was self-employed are more likely to be self-employed and less likely to work as an employee either in the farm or nonfarm sectors. If father was an employee, an adult child is more likely to be a nonfarm employee as well. Mother's self-employment is positively related to her children's choice of nonfarm self-employment and negatively related to their likelihood of becoming a farm worker.

Parents' educational attainment generally correlates strongly with individual's sector of employment increasing(decreasing) an adult child's probability of working in the nonfarm (farm) sector. University graduate parents, especially the mother, sharply increase the likelihood that an individual works in the nonfarm sector.

The value of owned agricultural land from the past 10 years, representing a farm sector specific-capital endowments, statistically significantly increases the likelihood of farm self-employment. Intuitively, schooling is positively (negatively) associated with the likelihood of working in the nonfarm (farm) sector. A high school degree strongly increases the probability of non-farm employment. A university degree has an especially large, positive effect on non-farm employment, but statistically significantly only for women. The effect of age is nonlinear, but statistically significant only for the nonfarm sector, concave (convex) for self-employment (employee). The age of the maximum predicted probability for non-farm self-employment is 52 whereas the maximum predicted probability for non-farm employee is reached at the age of 20 and is getting lower probability when approaching 70 years old. Younger individuals are likely to work as nonfarm employees while older people, with more experience, have a higher likelihood to operate their own self-employed business. Women are more likely to operate a nonfarm self-employed business rather than a farm. As expected, rural individuals work disproportionately in the farm sector whereas urban individuals mainly work in the nonfarm sector.

Table 4.3: Average marginal effects for multinomial logit

	Farm worker		Farm self-employed		NF employee		NF self-employed	
	Coef.	Std. err.	Coef.	Std. err.	Coef.	Std. err.	Coef.	Std. err.
Mother SE	-0.025*	0.013	-0.009	0.013	-0.006	0.019	0.039**	0.017
Mother worker	-0.003	0.011	-0.046**	0.019	0.060***	0.021	-0.011	0.025
Father SE	-0.044**	0.019	0.021	0.020	-0.029	0.030	0.052*	0.029
Father worker	-0.044***	0.016	-0.001	0.022	0.057**	0.024	-0.012	0.026
Mother elementary	-0.016	0.012	-0.024	0.015	0.009	0.021	0.031	0.022
Mother Jr.high sch	-0.031	0.030	-0.146**	0.064	0.077	0.052	0.101**	0.048
Mother Sr. high sch	-0.034	0.033	-0.060	0.050	-0.054	0.057	0.147**	0.058
Mother University	0.222***	0.082	-1.018***	0.068	0.247**	0.108	0.549***	0.099
Father elementary	-0.001	0.012	-0.011	0.017	0.039**	0.019	-0.027	0.024
Father Jr.high sch	-0.058**	0.029	0.058*	0.033	-0.009	0.036	0.009	0.039
Father Sr. high sch	-0.016	0.027	-0.053	0.034	0.038	0.040	0.030	0.040
Father University	-0.098	0.088	-0.061	0.094	0.026	0.077	0.134*	0.075
Value of agri land	-0.000	0.000	0.0001***	0.00004	0.000	0.000	-0.000	0.000
Years of schooling	-0.014***	0.005	-0.016***	0.005	0.010	0.009	0.021**	0.009
Female*years school	0.010	0.009	0.012	0.008	-0.006	0.017	-0.016	0.014
Completed grade 6	0.029	0.025	0.042*	0.022	0.009	0.045	-0.081*	0.045
Completed grade 9	-0.009	0.021	0.030	0.022	0.009	0.035	-0.030	0.037
Completed grade 12	0.015	0.021	0.002	0.024	0.070*	0.037	-0.087**	0.038
Completed grade 16	-0.030	0.048	0.027	0.053	0.050	0.054	-0.047	0.054
Female*Dcom6	-0.064	0.045	-0.058	0.039	0.066	0.072	0.056	0.071
Female*Dcom9	-0.064	0.046	-0.005	0.048	-0.009	0.071	0.078	0.068
Female*Dcom12	0.0005	0.043	0.012	0.058	0.041	0.070	-0.054	0.069
Female*Dcom16	0.024	0.072	-0.152*	0.087	0.174*	0.092	-0.046	0.086
age	-0.002	0.002	0.003	0.002	-0.023***	0.004	0.022***	0.003
age^2	0.000	0.000	0.000	0.000	0.0002***	0.00005	-0.0002***	0.00004
Female	-0.036	0.024	-0.117***	0.024	-0.038	0.063	0.192***	0.055
HH head	-0.046***	0.012	0.070***	0.015	0.039*	0.023	-0.062**	0.024
Durban	-0.038***	0.011	-0.149***	0.013	0.123***	0.020	0.064***	0.017
Joint test for parents' information	(base case)		169.69***		27.39**		26.68**	
Pseudo R^2	0.2788							
N. of obs	6,486							

***P<0.01, **P<0.05, and *P<0.10. Standard errors are clustered at community level and attrition weights are applied. Provincial dummy variables are included but not shown.

OLS coefficient estimates of earnings equation (2), assuming exogenous sectors of employment, are reported in Appendix Table B1. We report the estimated average effects of schooling, including sheepskin effects, in Table 4.4 by calculating estimated average log earnings changes from an additional year of schooling for each gender – occupation cohort, evaluated at the means of the other control variables. The results in column (1) suggest that, on average, females gain more from schooling than males do. Earnings gains from schooling are likewise greater in the nonfarm sector than in the farm sector; nonfarm employees enjoy the highest statistically significant marginal earnings gains from schooling, for both males (10.6%) and females (36.7%).

Table 4.4: Average effects of schooling, including sheepskin effects, on earnings from OLS and 2-stage estimation correcting for selection in the first stage

	OLS (1)	Selection (2)
Male farm worker	-0.091	-0.184
Male farm self-employed	-0.014	0.024*
Male NF worker	0.106***	0.144***
Male NF self-employed	0.078*	0.080*
Female farm worker	0.316**	0.019
Female farm self-employed	0.264*	0.026
Female NF worker	0.367**	0.154***
Female NF self-employed	0.344	0.100*
No. of observations	6,486	6,486

***P<0.01, **P<0.05, and *P<0.10.

Table 4.5 reports results from the second stage selection models based on the first stage multinomial logit model in Table 4.3. Not all selection terms are individually statistically significant, but there exist selection effects in each sector and these selection terms are statistically jointly statistically significant only for farm and non-farm worker equations. Both negative and positive correlations of the selection terms exist in all equations, implying that unobserved parts relating to the probabilities of selection into that sector could bias the estimated marginal earnings of education in the absence of this two-stage selection estimation strategy.[32]

The estimated coefficients on years of schooling for male vary from -0.095 to 0.148, with the highest estimates consistently associated with nonfarm employees. High school and college degrees are positively associated with earnings from non-farm occupations. Although coefficients on schooling variables are not statistically significant for all sectors, they are statistically jointly significant, except in the farm self-employment equation. Other coefficients on individual characteristics are statistically significant and show the expected relationship with log earnings: life cycle effects, lower earnings for female, and higher earnings for household heads if they are in the nonfarm sector. On average, the highest log earnings occur at the estimated age of 67 years old for farm workers, 58 years old for farm self-employed, and 43 years old for the non-farm occupations. If farm workers had agricultural land, each additional one million rupiah of the land value is associated with an 0.4 percent

[32] Exclusion restriction tests are shown in Appendix Table C1. Although the joint test for all instruments are statistically significant in determining the earnings equations, mother's occupation variables are always statistically insignificant. We have at least a set of variables that is valid for being instruments.

increase in earnings. However, the equivalent agricultural land effects are tiny – only 0.01 percent increase in earnings – for the farm self-employed.

We also summarize the average effects of schooling on wage differentials after correcting for selection in the rightmost column of Table 4.4 as estimated average differences in log earnings. Firstly, there are considerable differences between the OLS estimates and the selection correction estimates, potentially bias introduced from failure to control for selection into sector of employment. The marginal earnings gains from schooling are substantial higher for females in the OLS model while they are approximately in the same range between males and females in the selection model. Secondly, both models show a sharp intersectoral difference in the marginal earnings associated with education, with higher gains in the non-farm sector than in the farm sector. In the selection model, the marginal earnings gains from schooling for individuals in the non-farm sector vary from 8.0-15.4 percent while those in the farm sector are approximately around zero. Thirdly, there are also significant signaling effects, as reflected in higher marginal earnings associated with education for non-farm workers than non-farm self-employed (14.4 percent compared to 8.0 percent for males, and 15.4 percent compared to 10.0 percent for females). Lastly, gender differences are comparatively modest. This is likely because men enjoy higher marginal earnings from education up through junior high school, and with sheepskin effects women start to enjoy higher marginal earnings from junior high school.

In addition, we graphically illustrate these results combining education and sheepskin effects plus the differences for three-way comparisons between farm-nonfarm sector, self-employed/employee, and gender, varying years of schooling, sector of employment and gender, given the means of other control variables in figures 4.3-4.5. Figure 4.3 compares between workers and the self-employed, conditional on sector, so as to explore the prospective signaling effects of education. When we correct for selection into sector of employment, the signaling effect is striking for female graduating from junior high school and working in the non-farm sector. With no schooling or low levels of schooling, non-farm self-employed men earn more than non-farm workers. With a higher rate of marginal earnings gains from schooling (14.4 compared to 8.0 percent), male non-farm workers' earnings increase considerably with higher levels of education and eventually catch up their self-employed earnings after graduating from college.

Figure 4.4 compares between farm and nonfarm sectors, given an individual working as a worker or as a self-employed business owner. The non-farm sector yields higher expected earnings after primary school completion for workers, and over the full range of schooling years for the self-employed, with both higher gains for each extra year of schooling and more positive sheepskin effects. The differences in log earnings between the non-farm and the farm self-employed are statistically significant. For workers, these higher gains are statistically significant after graduating from primary school, reflecting more professional training required in the formal job markets.

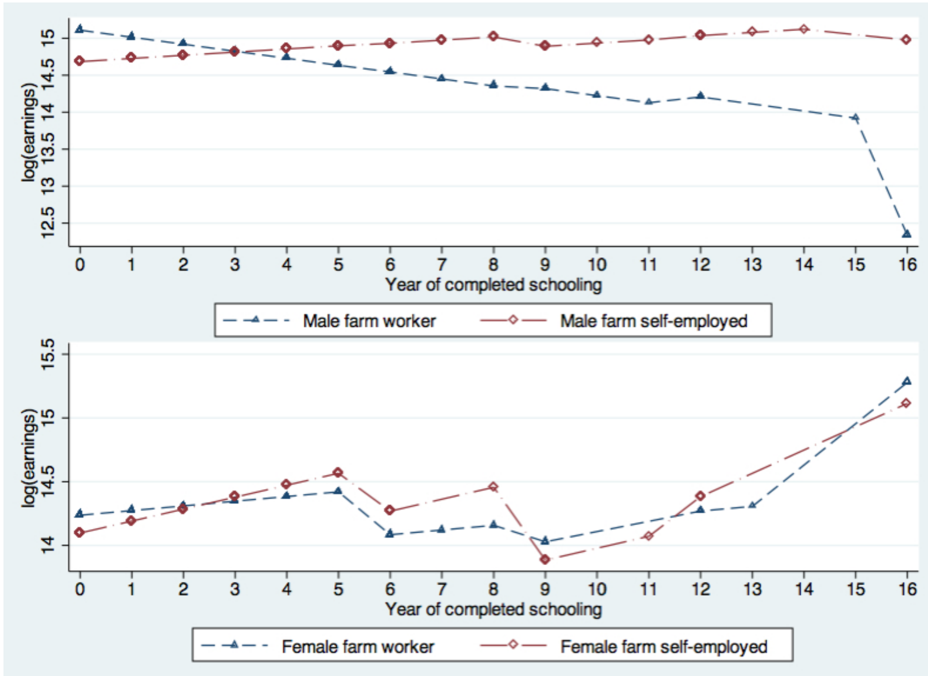
Figure 4.5 compares between genders. Male-female differences in log earnings consistently favor men among the nonfarm self-employed, but otherwise favor men only at lower levels of educational attainment and women at higher. The graphs show a crossover point after graduating from senior high school (grade 12) for farm workers, from college for farm self-employed individuals, and from junior high school (grade 9) for non-farm workers. This suggests that sheepskin effects play a larger role for females in higher levels of schooling attainment, and that secondary and tertiary education have a more pronounced effect on women's labor market outcomes than on men's.

The log earnings differences between the non-farm and the farm sectors, and between males and females, are more pronounced and mostly statistically significant in the selection corrected models, as compared to the OLS estimates. This could indicate that correcting for occupational selection isolates more clearly the differences in gains across sectors and between the sexes. Education, meanwhile, seems to play a more prominent role in sorting workers among sectors than in determining earnings within a sector through signaling effects.

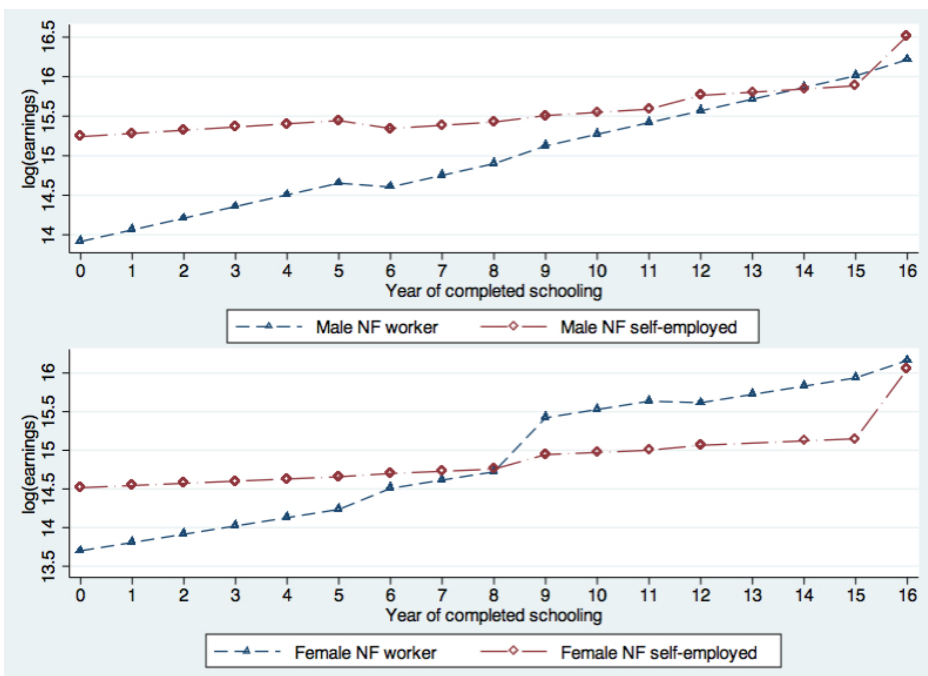
Table 4.5: Earning equations correcting for selection on employment sector (Eq. 4.5)

Dependent variable: log earnings	Farm wkr	Farm SE	NF wkr	NF SE
Years of schooling	-0.095	0.043	0.148***	0.041
	(0.059)	(0.050)	(0.036)	(0.044)
Fem* years schooling	0.132*	0.051	-0.041	-0.012
	(0.072)	(0.074)	(0.050)	(0.052)
Completed grade 6	0.006	-0.009	-0.197	-0.140
	(0.226)	(0.189)	(0.167)	(0.167)
Completed grade 9	0.059	-0.168	0.076	0.041
	(0.274)	(0.161)	(0.138)	(0.188)
Completed grade 12	0.172	0.017	0.001	0.133
	(0.243)	(0.219)	(0.131)	(0.192)
Completed grade 16	-1.489**	-0.232	0.057	0.593**
	(0.598)	(0.452)	(0.154)	(0.233)
Female*Completed grade 6	-0.379	-0.382	0.362	0.155
	(0.433)	(0.415)	(0.249)	(0.243)
Female*Completed grade 9	-0.225	-0.497	0.515**	0.118
	(0.435)	(0.423)	(0.252)	(0.263)
Female*Completed grade 12	-0.041	0.201	-0.127	-0.097
	(0.532)	(0.472)	(0.186)	(0.260)
Female*Completed grade 16	2.352***	0.588	0.061	0.281
	(0.879)	(1.076)	(0.227)	(0.369)
Age	0.134***	0.116***	0.172***	0.086***
	(0.035)	(0.029)	(0.015)	(0.023)
Age2	-0.001***	-0.001***	-0.002***	-0.001***
	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Female	-0.872***	-0.587**	-0.214	-0.725***
	(0.284)	(0.292)	(0.212)	(0.196)
Head	-0.049	-0.479	0.296***	0.210**
	(0.218)	(0.298)	(0.067)	(0.100)
Agricultural land value	0.004***	0.0001	0.00001	0.001*
	(0.001)	(0.0004)	(0.0002)	(0.0003)
Urban	-0.404	0.052	0.291***	0.212
	(0.272)	(0.297)	(0.091)	(0.134)
Selection terms:				
- Farm worker	0.159	0.889	-1.338***	-0.090
	(0.117)	(0.820)	(0.291)	(0.461)
- Farm self-employed	-0.456	-0.091	-0.428	-0.246
	(1.065)	(0.181)	(0.336)	(0.459)
- NF worker	-2.449**	-0.757	-0.121	0.508
	(1.013)	(0.715)	(0.120)	(0.489)
- NF self-employed	0.708	0.925	0.356	-0.256**
	(0.762)	(0.674)	(0.310)	(0.112)
Constant	10.629***	13.252***	10.693***	14.117***
	(1.276)	(1.086)	(0.347)	(0.896)
Joint test for schooling variables	15.85**	8.16	191.91***	23.46***
Joint test for female completed degrees	10.75**	2.78	8.92*	1.04
Adjusted R^2	0.22	0.15	0.34	0.17
N	550	1,190	3,012	1,734

***P<0.01, **P<0.05, and *P<0.10. Standard errors are clustered at community level and attrition weights are applied. Provincial dummy variables are included but not shown.

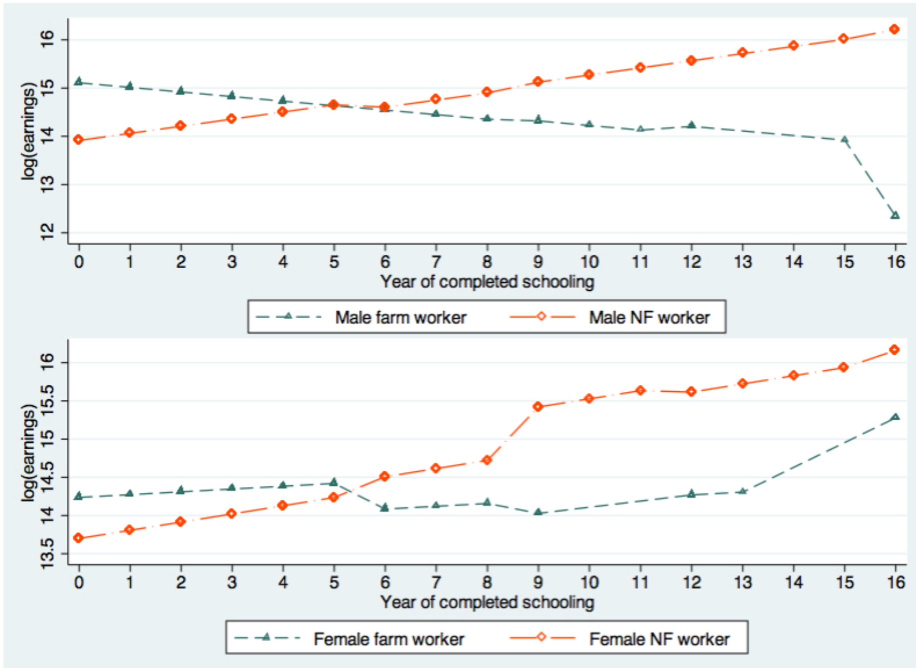


(a) Farm sector: worker vs. self-employed

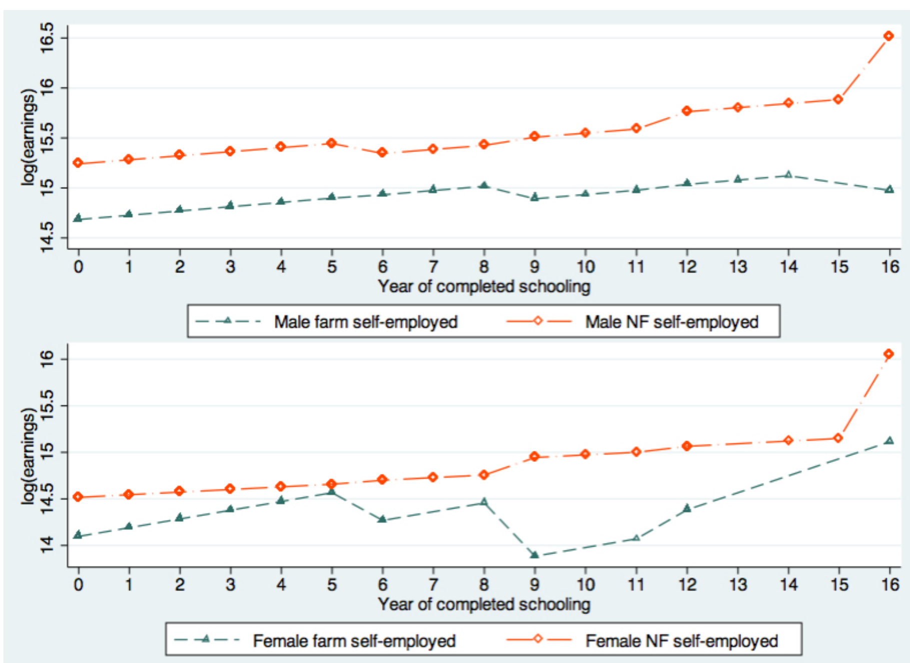


(b) Non-farm sector: worker vs. self-employed

Figure 4.3: Uncovering signaling effects: estimated log earnings by years of schooling, gender, and farm/non-farm sector, correcting for selection into sectors of employment

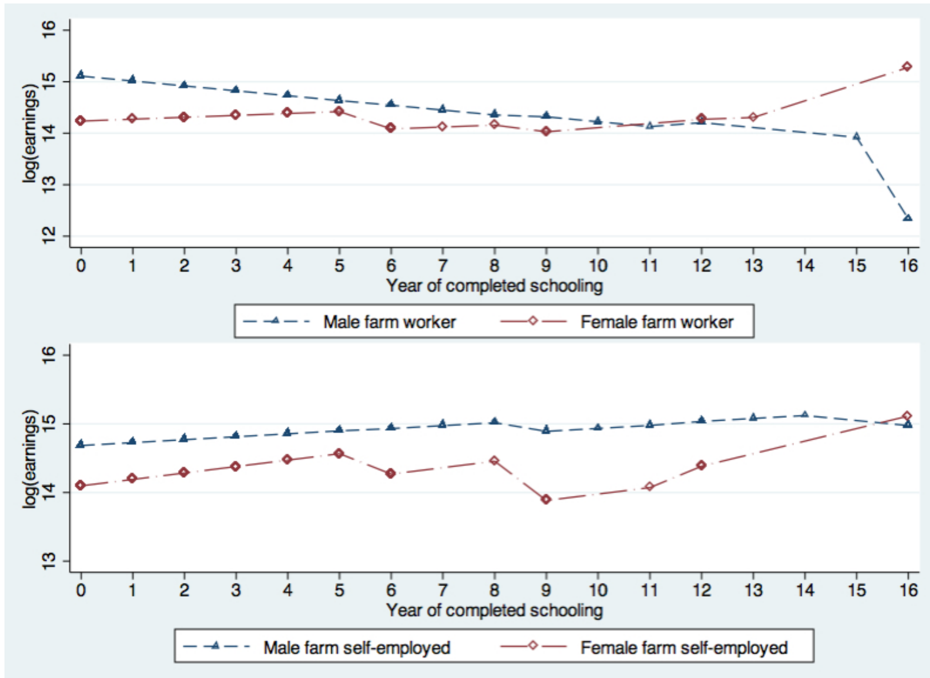


(a) Worker: farm vs non-farm sectors

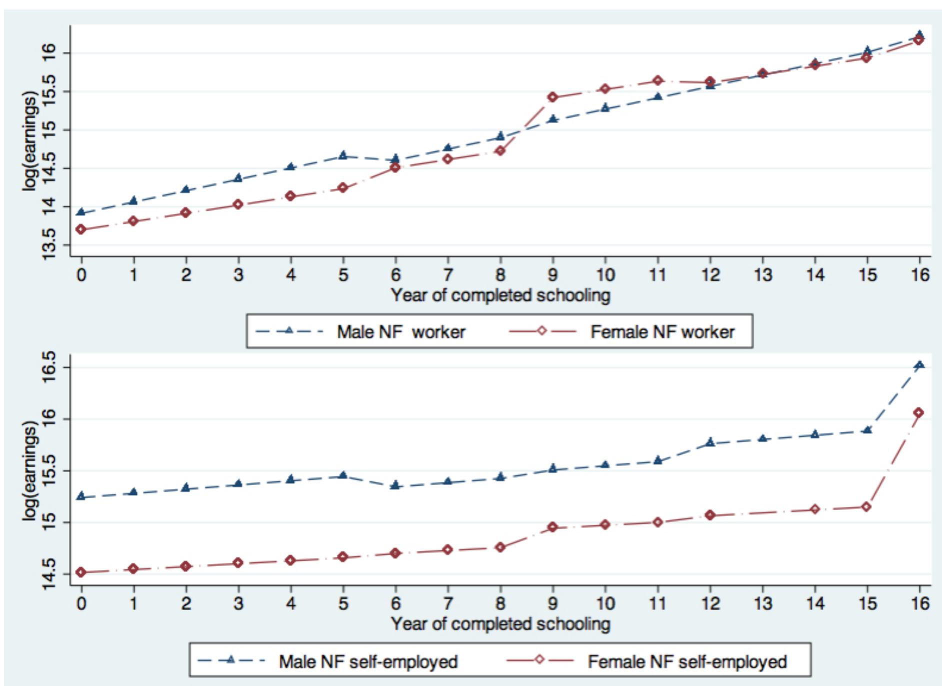


(b) Self-employed: farm vs. nonfarm sectors

Figure 4.4: Uncovering intersectoral differences: estimated log earnings by years of schooling across sectors and gender, given occupation, correcting for selection into sectors of employment



(a) Farm: male vs. female



(b) Nonfarm: male vs. female

Figure 4.5: Uncovering gender differences: estimated log earnings by years of schooling, gender, and sector of employment, correcting for selection into sectors of employment

4.5 Conclusions

Educational attainment can influence individual earnings through multiple pathways, especially by impacting marginal productivity or by signaling unobservable characteristics to prospective employers. These effects may differ between the farm and non-farm sectors and between men and women if labor markets are not fully integrated, as is likely the case in developing countries such as Indonesia. This paper explores the possibility that the marginal earnings gains associated with educational attainment vary significantly between the farm and non-farm sectors, between men and women, and between the employed and the self-employed.

We find that selection into sector of occupation – farm self-employed, farm worker, non-farm self-employed, or non-farm employee – is strongly affected by parental occupation and educational attainment, in ways that are quite intuitive. Moreover, failing to correct for these selection effects leads to considerable bias in the estimated marginal earnings gains associated with added educational attainment.

Once one corrects for selection into sector of occupation, we find sharp intersectoral differences in the marginal earnings gains associated with education, strongly favoring those in the nonfarm sector, and much more modest gender differences in the marginal earnings gains from schooling, seemingly because men tend to enjoy higher marginal earnings from educational up through junior high school, and women enjoy higher marginal earnings thereafter. There also appear to be important job market signaling effects, as reflected in the difference between the marginal earnings enjoyed by employees versus otherwise identical self-employed individuals. And significant sheepskin effects emerge in secondary and tertiary education, consistent with the signaling hypothesis.

These findings carry important implications for understanding patterns of private investment in education. If parents who expect their children to take over the family farm perceive that the earnings gains from education in farming are modest or negligible and they understand that the child will not need education as a signal of ability if she is to be self-employed, then they are likely to invest less in schooling for that child than would parents who expect their child to pursue a non-farm career as a salaried civil servant or firm employee, other things equal. Observed significant differences in the marginal earnings gains associated with educational attainment might help explain heterogeneous educational investment patterns in equilibrium.

REFERENCES

Barrett, C.B., M.R. Carter, and P. Timmer. 2010. "A Century-Long Perspective on Agricultural Development", *American Journal of Agricultural Economics* 92(2): 447-468.

Becker, G. S, and N. Tomes. 1979. "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility Nigel Tomes", *Journal of Political Economy* 87:1153–1189.

Belman, D. and J.S. Heywood. 1991. "Sheepskin Effects in the Returns to Education: An Examination of Women and Minorities", *The Review of Economics and Statistics* 73(4): 720-724.

Bjorklund A. and K. Salvanes. 2010. "Education and Family Background: Mechanisms and Policies", Chapter 3 In: Hanushek, E.A., Machin, S., and Woessmann, L. (Eds.), *Handbook of the Economics of Education Vol. 3*, Elsevier.

Black, S.E. and P.J. Devereux. 2010. "Recent Developments In Intergenerational Mobility", NBER Working Paper 15889.

Bourguignon, F., M. Fournier, and M. Gurgand. 2007. "Selection bias corrections based on the multinomial logit model: Monte Carlo comparisons." *Journal of Economic Surveys* 2(1): 174-205.

Card, D. 1999. "The Causal Effect of Education on Earnings," in: Ashenfelter, O. and Card, D. (Eds), *Handbook of Labor Economics, Vol. 3A*, Amsterdam: North-Holland.

Card, D. 2001. "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica* 69(5):1127–1160.

Carneiro, Pedro, James J Heckman, and Edward Vytlacil. 2001. "Estimating The Return to Education When It Varies Among Individuals." Working paper.

Carneiro, Pedro, James J Heckman, and Edward J Vytlacil. 2011. "Estimating Marginal Returns to Education." *American Economic Review* 101(October): 2754–2781.

Carneiro, Pedro, Michael Lokshin, Cristobal Ridao-cano, and Nithin Umapathi. 2011. "Average and Marginal Returns to Upper Secondary Schooling in Indonesia." World Bank Policy Research Working Paper 5878.

Comola, M. and L. Mello. 2013. "Salaried Employment and Earnings in Indonesia: New Evidence on the Selection Bias", *Applied Economics* 45: 2808-2816.

Davidson, R., and JY. Duclos, , 2000. "Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality", *Econometrica* 68 (6): 1435-1464.

Dubin, J. A. & McFadden, D. L. 1984. "An econometric analysis of residential electric appliance holdings and consumption.". *Econometrica, 52*(2), 345-362.

Duflo, E. 2001. "Schooling and Labor Market Consequences of School Construction in Indonesia : Evidence from an Unusual Policy Experiment", *American Economic Review* 91:795–813.

Fields, G. 1980. "Education and Income Distribution in Developing Countries: A Review of the Literature." World Bank working paper.

Flores-Lagunes, A. and A. Light. 2010. "Interpreting Degree Effects in the Returns to Education", *Journal of Human Resources* 45(2): 439-467.

Granado, F. Javier Arze del, Wolfgange Fengler, Andy Ragatz, and Elif Yanuz. 2007. “Investing in Indonesia’s Education: Allocation, Equity, and Efficiency of Public Expenditures”, World Bank Policy research working paper 4329.

Heckman, J., S. Urzua, and E. Vytlacil. 2006. “Understanding Instrumental Variables in Models with Essential Heterogeneity”, *Review of Economics and Statistics* 88:389–432.

Hellerstein, Judith and Melinda Morrill. 2011. “Dads and daughters: The changing impact of fathers on women’s occupational choices”, *Journal of Human Resources* 46: 333-372.

Huffman, W.E. and P.F. Orazem. 2007. “Agriculture and Human Capital in Economic Growth: Farmers, Schooling and Nutrition”, Chapter 43 in: Evenson, R. and Pingali, P. (Eds.) , Handbook of Agricultural Economics Vol. 3., Elsevier.

Hungerford, T. and G. Solon. 1987. “Sheepskin Effects in the Returns to Education”, *Review of Economics and Statistics* 69(1): 175-177.

Jaeger, D. A., and Page, M. E. 1996. “Degrees Matter: New Evidence on Sheepskin Effects in the Returns to Education”, *Review of Economics and Statistics 78*(4): 733–740.

Jensen, R. 2010. “The (Perceived) Returns to Education and the Demand for Schooling”, *Quarterly Journal of Econometrics* 125(2): 515-548.

Kramarz, F. and O.N. Skans. 2011. “With a Little Help from my Parents? Family Networks and Youth Labor Market Entry”, Working paper, Center for Research in Economics and Statistics (CREST).

Lassibille,G. and J. Tan. 2005. “The Returns to Education in Rwanda.” *Journal of African Economies* 14(1): 92–116.

Lewis, A. 1954. “Economic Development with Unlimited Supplies of Labor”, *The Manchester School* 22: 139-191.

Mincer, J. 1974. *Schooling Experience and Earnings*. NBER, New York.

Magruder, J.R. 2010. “Intergenerational Networks, Unemployment, and Persistent Inequality in South Africa”, *American Economic Journal: Applied Economics* 2(1): 62-85.

Murnane, R.J., Willett, J.B. and Levy,F. 1995. “The Growing Importance of Cognitive Skills in Wage Determination”, NBER Working Paper No. 5076.

Nguyen, T. 2008. “Information, Role Models and Perceived Returns to Education: Experimental Evidence from Madagascar”, MIT working paper.

Oyelere, R. U. 2010. “Africa’s education enigma? The Nigerian story”, *Journal of Development Economics 91*(1): 128–139.

Riley, J. G. 1979. “Testing the Educational Screening Hypothesis”, *Journal of Political Economy* 87:S227–S252.

Schultz, T. 1964. *Transforming Traditional Agriculture*, Yale University Press, New Heaven.

Siphambe, H. K. 2000. “Rates of return to education in Botswana.” *Economics of Education Review* 19:291–300.

Spence, M. 1973. “Job Market Signaling,” *The Quarterly Journal of Economics* 87(3): 355-374.

Strauss, J., K. Beegle, B. Sikoki, A. Dwiyanto, Y. Herawati and F. Witoelar. 2004. “The Third Wave of the Indonesia Family Life Survey (IFLS3): Overview and Field Report”. March 2004. WR-144/1-NIA/NICHD.

Strauss, J., F. Witoelar, B. Sikoki and AM Wattie. 2009. “The Fourth Wave of the Indonesia Family Life Survey (IFLS4): Overview and Field Report”. March 2009. WR-675/1-NIA/NICHD.

Suryadarma, D. and A. Suryahadi. 2010. “Determinants of Education Attainment in Developing Countries: Can Higher Skills Compensate for Poverty?” RAND working paper.

Suryadarma, D., A. Suryahadi and S. Sumarto. 2006. “Causes of Low Secondary School Enrollment in Indonesia” SMERY Research Institute working paper.

Timmer, P. 2009. “A World without Agriculture: The Structural Transformation in Historical Perspective”, Henry Wendt Lecture, American Enterprise Institute, 2008.

Wolpin, K. 1977. "Education and Screening", *American Economic Review* 67: 949-58.

APPENDIX 4A

ATTRITION AND SELECTION ISSUES

Although IFLS is well-known for its high overall participation rate across survey rounds, when combining specific survey modules together across years, many missing values arise, resulting in attrition from the estimation sample. Some survey questions are not applicable to all individuals, hence missing values. Moreover, in our study, we focus on employed individuals whose earnings and schooling are reported, resulting in sample selection problem. Moffitt, Fitzgerald, and Gottschalk (1999, hereafter MFG), and Wooldridge (2002) suggest using inverse probability weighting (IPW) to deal with attrition and selection problems in a more general model. The IPW approach, it is assumed selection on observables such that, there is a set of auxiliary variables, $\boldsymbol{z}$, affecting attrition and selection propensities that is related to the density of a dependent variable of interest, y, conditional on $\boldsymbol{x}$.

As IFLS can be analyzed as cross-sectional data, we use IFLS4 (year 2007) earnings values as our main dependent variable and only track back to the earlier survey waves for instrumental variables. We then treat IFLS4 observations as the baseline, (y_{i1}, $\boldsymbol{x}_{i1}$), $A_i = 0$ if we have schooling-earnings data and match with parental data and/or agricultural land data from IFLS1-2 (non-missing data in all variables), and $A_i = 1$ otherwise (including both attrition and selection). Table 1A presents testing results from the attrition probit regression for the full sample and the estimation subsample, as in Table 1. For the full sample, attrition exists for the log earnings and reported years of schooling that we use for this study. The auxiliary variables include log household asset value, attrition rate at community level, dummy for existing in all four waves, dummy for demography variables, including Islam race, age and age squared, female, household head, and household size, and geography variables, including urban and provincial dummies, all in 2007. The pseudo R^2 statistic can be interpreted as the proportion of attrition that is non-random. Wald statistics are then performed to test whether observables jointly explain the predicted attrition probability. As we have high pseudo R^2 and all Wald tests are statistically significant, attrition appear related to household, demographic, and geographic characteristics.

Table A1: Attrition probit regressions

	Full sample (1)	Estimation sample (2)
Pseudo R2	0.2270	0.1929
Wald test (p-value report)		
log(household asset)	0.000	0.000
Village attrition rate	0.000	0.000
Exiting in all 4 waves	0.000	0.000
Demography[a]	0.000	0.000
Geography[b]	0.000	0.000
No. of observations	28,047	28,047

Columns (1) and (2) are according to samples 1 and 2 as reported in Table 1. [a]Demographic variables include household size, dummy for islam race, age and age squared, female, and household head. [b]Geographic variables include urban and provincial dummies. Wald tests for demography and geography variables are joint tests. Standard errors are clustered at community level and survey weights correcting for survey attrition are applied.

Following MFG and Baulch and Quisumbing (2011), we use the inverse probability weighted method to correct for possible biases. We create the ratio of predicted values from the restricted regression and unrestricted regression of the reversed attrition probit, where the dependent variable, RA = 1 for nonattrited observations. The unrestricted regression includes the same explanatory variables as the attrition probit regressions, while the restricted regression only includes log household asset, household size, and dummies for urban and provincial variables. We create inverse probability weights specifically for each sample group as each sample has a different set of non-missing observations. The IPW vary from 0.15 to 87.19 with mean 1.01 for sample 1, and from 0.09 to 32.47 with mean 1.02 for sample 2.

Table A2 compares the OLS regressions between the model using inverse probability weights and the model using survey weights. Their coefficients are statistically insignificantly different from each other.

Table A2: OLS estimates of log earnings equation with interactions among schooling, gender, and occupation

OLS log(earnings)	(1) Attrition weights	(2) Survey weights
Years of schooling	0.036*	0.043**
	(0.019)	(0.017)
Fem*yearsch07	0.068***	0.042
	(0.023)	(0.027)
Farm self-employed	0.046	0.069
	(0.128)	(0.113)
NF employee	0.009	0.235**
	(0.130)	(0.106)
NF self-employed	0.509***	0.573***
	(0.121)	(0.104)
Farm SE *fem	0.277	0.173
	(0.217)	(0.229)
NF wkr *fem	0.453**	0.290
	(0.186)	(0.197)
NF SE *fem	0.402**	0.394**
	(0.185)	(0.192)
Farm SE *yearsch	-0.033**	-0.029*
	(0.017)	(0.015)
NF wkr *yearsch	0.020	0.013
	(0.015)	(0.013)
NF SE *yearsch	-0.027*	-0.026**
	(0.015)	(0.013)
Farm SE*fem*sch	-0.051	-0.025
	(0.037)	(0.037)
NF wkr*fem*sch	-0.042*	-0.011
	(0.024)	(0.028)
NF SE*fem*sch	-0.071***	-0.049*
	(0.025)	(0.028)

OLS log(earnings)	(1) Attrition weights	(2) Survey weights
Completed grade 6	0.149**	0.086
	(0.071)	(0.064)
Completed grade 9	0.183***	0.143**
	(0.066)	(0.056)
Completed grade 12	0.181***	0.191***
	(0.060)	(0.055)
Completed grade 16	0.323***	0.377***
	(0.073)	(0.067)
Age 07	0.116***	0.122***
	(0.007)	(0.007)
Age^2 07	-0.001***	-0.001***
	(0.0001)	(0.0001)
Female 07	-0.928***	-0.869***
	(0.166)	(0.176)
Head 07	0.163***	0.187***
	(0.044)	(0.038)
Durban 07	0.129***	0.130***
	(0.045)	(0.046)
Constant	12.209***	11.877***
	(0.191)	(0.197)
Adjusted R^2	0.28	0.29
N. of obs	12,675	12,675

***P<0.01, **P<0.05, and *P<0.10. Standard errors are clustered at community level. Other variables not showing are dummies for provinces.

Table B1: OLS estimations of full models (with all interactions)

	(1) Farm wkr		(2) Farm SE		(3) NF worker		(4) NF SE	
	Sample1	Sample2	Sample1	Sample2	Sample1	Sample2	Sample1	Sample
constant	12.102***	12.222***	0.074	0.133	0.189	-0.231	0.462***	0.318
	(0.216)	(0.318)	(0.177)	(0.390)	(0.174)	(0.253)	(0.153)	(0.202)
yearsch	0.083*	0.129	-0.055	-0.181	-0.058	-0.057	0.002	-0.096
	(0.043)	(0.101)	(0.057)	(0.140)	(0.054)	(0.113)	(0.051)	(0.108)
Female_07#c.yearsch	0.042	-0.004	-0.028	0.153	-0.084	0.050	-0.118*	-0.039
	(0.059)	(0.124)	(0.090)	(0.178)	(0.079)	(0.153)	(0.071)	(0.139)
Dcomgrad6	-0.122	-0.730	0.169	0.856	0.385	0.688	-0.081	0.519
	(0.224)	(0.598)	(0.265)	(0.698)	(0.266)	(0.648)	(0.248)	(0.609)
Dcomgrad9	0.161	0.100	-0.060	0.122	-0.012	-0.040	-0.281	-0.004
	(0.213)	(0.380)	(0.261)	(0.471)	(0.242)	(0.426)	(0.248)	(0.456)
Dcomgrad12	-0.014	-0.590	0.138	0.918*	0.388*	0.757*	-0.012	0.785*
	(0.196)	(0.379)	(0.268)	(0.525)	(0.229)	(0.406)	(0.230)	(0.430)
Dcomgrad16	-0.688	-1.854***	0.526	1.621*	1.145*	2.213***	0.944	2.490*
	(0.631)	(0.693)	(0.762)	(0.918)	(0.649)	(0.736)	(0.652)	(0.705)
Female	-0.851***	-1.093***	0.169	0.080	0.410	0.386	0.440**	0.432
	(0.190)	(0.208)	(0.257)	(0.489)	(0.257)	(0.423)	(0.216)	(0.264)
Female#Dcomgrad6	-0.007	0.337	0.390	-1.168	0.253	-0.512	0.369	0.007
	(0.324)	(0.783)	(0.475)	(0.927)	(0.392)	(0.903)	(0.363)	(0.830)
Female#Dcomgrad9	0.224	0.617	-0.942**	-0.894	0.364	-0.205	0.007	-0.354
	(0.343)	(0.578)	(0.459)	(0.734)	(0.402)	(0.698)	(0.363)	(0.681)
Female#Dcomgrad12	-0.134	0.629	0.136	-0.484	0.114	-0.902	0.319	-0.510
	(0.501)	(0.557)	(0.699)	(0.843)	(0.514)	(0.656)	(0.556)	(0.692)
Female#Dcomgrad16	1.432*	2.125***	-0.500	-1.864	-1.284	-2.350***	-1.229	-1.960*
	(0.801)	(0.800)	(1.194)	(1.311)	(0.862)	(0.901)	(0.825)	(0.842)

	Sample1	Sample2
Age	0.117***	0.145***
	(0.007)	(0.012)
Age^2	-0.001***	-0.002***
	(0.0001)	(0.0001)
Head	0.156***	0.056
	(0.043)	(0.071)
Agricultural land value		0.0002**
		(0.0001)
Urban	0.128***	0.122*
	(0.045)	(0.067)
R2_A	0.28	0.29
N	12,675	6,486

***P<0.01, **P<0.05, and *P<0.10. Standard errors are clustered at community level and attrition weights are applied. Coefficients on columns (2)-(4) are seen as incremental magnitude from the base case, farm worker. Coefficients on age, age^2, head, and urban dummy are for the whole model.

Table C1: Exclusion restriction tests

	(1) F/p-value	(2) F/p-value
All	3.96***	3.42***
	0.0000	0.0000
Parents' occupation	4.40**	4.13***
	0.0017	0.0028
Mother's occupation	1.59	1.71
	0.2047	0.1818
Father's occupation	5.09***	5.41***
	0.0066	0.0048
Parents' education	3.11***	2.79***
	0.0008	0.0025
Mother's education	1.98*	1.77
	0.0817	0.1187
Father's education	1.53	2.16*
	0.1807	0.0577

***P<0.01, **P<0.05, and *P<0.10. Model (1) uses employment sector variables as solely dummy variables for each sector in the earning equations. Model (2) uses both employment sector dummy variables and their interactions with degree completion and gender variables, the same model as Table B1.